CERTIFICATE OF EXPRESS MAILING-37 C.F.R. § 1.10

"Express Mail" mailing label number: EL 452 904 740 US
Date of Deposit: March 23, 2001

US PATENT APPLICATION

INCREASED PRODUCTION OF SECRETED PROTEINS BY RECOMBINANT EUKARYOTIC CELLS

Number of Figures: Thirty-One (31) comprising Thirty-Five (35) Sheets

Inventors: Merja E. Penttila, a citizen of Finland, residing at
Vattuniemenkatu 22A2
Helsinki, Finland 00210

Michael Ward, a citizen of Great Britain, residing at
4372 24th Street
San Francisco, CA 94114

Huaming Wang, a citizen of the United States, residing at
4337 Calypso Terrace
Fremont, CA 94555

Mari J. Valkonen, a citizen of Finland, residing at
Harjutorinkatu 1A16
Helsinki, Finland 00500

Markku L.A. Saloheimo, a citizen of Finland, residing at
Kyntäjäntie 18b
Helsinki, Finland 00390

Assignee: **Genencor International, Inc.**
925 Page Mill Road
Palo Alto, CA 94304-1013
ph: (650) 846-7500

Docket No. GC590-2

PATENT GC590-2

INCREASED PRODUCTION OF SECRETED PROTEINS BY RECOMBINANT EUKARYOTIC CELLS

Field of the Invention

This invention relates to cells which have been genetically manipulated to have an elevated unfolded protein response (UPR) resulting in an increased capacity to produce secreted proteins.

Related Applications

This application is a Continuation-In-Part of United States Application No. 09/534,692, filed March 24, 2000, and is hereby incorporated by reference in its entirety.

Background of the Invention

The secretory pathway of eukaryotic organisms is of interest since cells can be engineered to secrete a particular protein of interest. The secretory pathway starts by translocation of the protein into the lumen of the endoplasmic reticulum (ER). In the ER the proteins fold into their final three-dimensional conformation and the core part of the N-glycans are attached to them. A quality control mechanism involving the proteins calnexin and calreticulin also resides in the ER, letting only completely folded proteins continue on the secretory pathway

to the next compartment (Hammond and Helenius, 1995, Curr. Opinion Cell Biol. 7:523-529). Secretory proteins that do not fold properly are transported back to the cytoplasm by the translocation machinery and are degraded by the proteasome system (Wiertz et al., 1996, Nature 384:432-438).

The folding and glycosylation of the secretory proteins in the ER is assisted by numerous ER-resident proteins. The chaperones like Bip (GRP78), GRP94 or yeast Lhs1p help the secretory protein to fold by binding to exposed hydrophobic regions in the unfolded states and preventing unfavourable interactions (Blond-Elguindi et al., 1993, Cell 75:717-728). The chaperones are also important for the translocation of the proteins through the ER membrane. The foldase proteins like protein disulphide isomerase and its homologs and prolyl-peptidyl cis-trans isomerase assist in formation of disulphide bridges and formation of the right conformation of the peptide chain adjacent to proline residues, respectively. A machinery including many protein components also resides in the ER for the addition of the N-linked core glycans to the secretory protein and for the initial trimming steps of the glycans.

The levels of the chaperone and foldase proteins found in the ER are regulated at the transcriptional level. For each gene there is a basic level of transcription that can be increased in response to various stimuli. A large amount of secretory protein in the ER (secretory load) can induce the mammalian GRP78 gene, and this induction is mediated through the NF-κB transcription factor (Pahl and Baeuerle,1995, EMBO J. 14:2580-2588). Furthermore, the ER chaperone and foldase genes are upregulated when the amount of unfolded protein increases in the ER. This induction has been named unfolded protein response (UPR) and it has been described in mammalian cells, yeast and filamentous fungi (McMillan et al., 1994, Curr. Opinion in Biotechnol. 5:540-545). The induction can be caused by treatment of cells with reducing agents like DTT, by inhibitors of core glycosylation like tunikamycin or by Ca-ionophores that deplete the ER calcium stores. The promoters of mammalian and yeast genes regulated by

UPR have a conserved sequence region called UPR element, where the transcription factor responsible for the induction binds.

When the unfolded protein response pathway is active, a signal is tranduced from the ER lumen to the transcription machinery in the nucleus. A protein implicated in the UPR induction is the IRE1 protein of yeast (Cox et al., 1993, Cell 73:1197-1206, Mori et al., 1993, Cell 74:143-156). It is large protein having a transmembrane segment anchoring the protein to the ER membrane. A segment of the IRE1 protein has homology to protein kinases and the C-terminal tail has some homology to RNAses. It is believed that the IRE1 protein may be the first component of the UPR signal transduction pathway, sensing the ER lumen for the presence of unfolded proteins and transmitting the signal eventually to a transcription factor inducing the ER-protein genes. It has been reported that the IRE1 protein oligomerizes and gets phosphorylated when the UPR is activated (Shamu and Walter, 1996, EMBO J. 15:3028-3039). Over-expression of the IRE1 gene in yeast leads to constitutive induction of the UPR (Id.). Phosphorylation of the IRE1 protein occurs at specific serine or threonine residues in the protein.

Another protein reportedly implicated in the regulation of the UPR pathway is PTC2, a yeast protein phosphatase encoded by the PTC2 gene (Welihinda et al., 1998, Mol. Cell. Biol. 18, 1967-1977). The IRE1 protein is phosphorylated when the UPR pathway is turned on (Shamu and Walter, 1996, EMBO J. 15:3928-3039), and PTC2 dephosphorylates the IRE1 protein and regulates the UPR.

It has further been reported that the yeast transcription factor mediating the UPR induction of the chaperone and foldase genes is the HAC1 protein (Cox and Walter, 1996, Cell 87:391-404, Sidrauski et al., 1996, Cell 87:405-413). It belongs to the bZIP family of transcription factors, having a basic DNA-binding region and a leucine zipper dimerisation domain. The binding of the HAC1 protein to the UPR element of ER-protein gene promoters has been

demonstrated (Mori et al., 1998, J. Biol. Chem. 273: 9912-9920). The action of the HAC1 protein is regulated by its amount in the cells; none of the protein can be found in uninduced cells and upon UPR induction it appears rapidly. The HAC1 protein amount is dependent of the splicing of the respective mRNA. In uninduced conditions the intron present in the HAC1 gene close to the translation termination codon is not spliced off, and this intron prevents the formation of HAC1 protein by preventing the translation of the mRNA (Chapman and Walter, 1997, Curr. Biol. 7, 850-859, Kawahara et al., 1997, Mol. Biol. Cell 8, 1845-1862). When UPR is induced, the intron is spliced and the mRNA is translated to form HAC1 protein that activates the promoters of its target genes. The HAC1 intron is spliced by an mechanism not currently described for any other system, involving the RNAse activity of the IRE1 protein and a tRNA ligase (Sidrauski and Walter, 1997, Cell 90, 1031-1039, Gonzales et al., 1999, EMBO J. 18, 3119-3132, Sidrauski et al., 1996, Cell 87, 405-413). The unfolded protein response can be induced constitutively in yeast by transformation with a UPR inducing version of the HAC1 gene. (Cox and Walter, supra.)

Thus, as indicated above, there are a number of reports regarding the secretory pathway. Additionally, there are reports on how to increase secretion so as to provide greater yields of heterologous proteins. Greater yields of protein are generally of interest to industry to provide more of a particular protein and to facilitate purification.

For example, in one report random mutagenesis of the host organism has been performed followed by screening for increased yield of a secreted protein. In another report, there has been fusion of a heterologous protein to an efficiently secreted endogenous protein in order to increase the yield of secretion of the heterologous protein. Both of these methods have been of limited success and other methods to improve protein secretion are desirable.

In other studies, there has reportedly been increased yields of secreted heterologous proteins in yeast by either over-expression or deletion of the yeast

ER foldase or chaperone genes on an individual or pairwise basis. For example, over-expression of either the protein disulphide isomerase (PDI) or the KAR2 (homologous to the gene for the mammalian ER chaperone BiP) genes in yeast has been shown to increase the extracellular accumulation of certain secreted heterologous proteins (Robinson et al., 1996, Bio/Technology, 12:381-384; Harmsen, et al., 1996, Appl. Microbiol. Biotechnol., 46:365-370). In contrast, deletion of the CNE1 gene, encoding an ER chaperone homologous to mammalian calnexin, reportedly can lead to increased secretion of a heterologous protein (Parlati et al., 1995, J. Biol. Chem. 270:244-253, Harmsen, supra.). The effect of over-expression or deletion of individual or pairs of ER chaperones or foldases has also been reported on in filamentous fungi, however, increased secretion was not obtained. (Punt, et al., 1998, Appl. Microbiol. Biotech, 50:447-454; Wang, et al., 2000, Current Genetics, 37:57-64).

Therefore, it is desirable to provide new methods to increase production of secreted proteins in eukaryotic cells which are simple and consistent. It is also desirable to provide compositions such as novel genes to be used in methods for the increased production of secreted proteins. It is further desirable to provide eukaryotic cells according to the invention which are transformed with heterologous genes so as to have an increased capacity to produce secreted proteins.

Summary of the Invention

Provided herein are methods for increasing the secretion of a heterologous protein in a cell comprising inducing an elevated unfolded protein response (UPR). The increase in protein secretion is compared to a level of protein secreted by the cell when the UPR is not elevated by the methods described herein. In one aspect, the method includes inducing the elevated UPR by increasing the presence of the HAC1 protein in the cell. In one aspect of the invention, the presence of the HAC1 protein can be increased by a number of methods. For example, the HAC1 gene can be overexpressed compared to its

native state. Overexpression can be achieved by a variety of ways including the use of preferred vectors and promoters as further described herein. In one embodiment, the HAC1 protein is increased in a cell by transformation of said cell by a nucleic acid comprising a UPR inducing form of a HAC1 recombinant nucleic acid.

The HAC1 nucleic acid encoding a HAC1 protein can be from a variety of sources. It is understood that in one embodiment, HAC1 is used interchangeably with hac1, hacA, etc., and one embodiment is meant to encompass HAC1 homologs. Additionally, the skilled artisan can ascertain by the context whether the HAC1 is a nucleic acid, protein or either. In one embodiment, a HAC1 nucleic acid is isolated from yeast. In another embodiment, a HAC1 nucleic acid is isolated from filamentous fungi such as *Trichoderma* or *Aspergillus*.

In another aspect of the invention, the elevated UPR is induced by modulating the levels of IRE1 protein or PTC2 protein in said cell. Nucleic acids encoding IRE1 or PTC2 can be isolated from yeast or filamentous fungi such as *Trichoderma* or *Aspergillus*. In a preferred embodiment the nucleic acid encoding IRE1 or PTC2 is isolated from *T. reesei, A. nidulans* or *A. niger*.

The cell from which the protein is secreted can be any cell having an UPR. Cells having an UPR include all eukaryotes including but not limited to mammalian cells, insect cells, yeast and filamentous fungi.

Also provided herein is an isolated nucleic acid encoding a HAC1 protein, wherein said HAC1 has unfolded protein response inducing activity and has less than 50% similarity to yeast HAC1 protein. In another embodiment, an isolated nucleic acid encoding a HAC1 protein is provided, wherein said HAC1 protein has unfolded protein response inducing activity and wherein said HAC1 comprises a DNA binding region that has greater than 70% similarity to the DNA binding region of filamentous fungi HAC1. Embodiments of a DNA binding region are shown at amino acids 84-147 of the *T. reesei* protein shown in Figure 10, at

amino acids 53-116 of the *A. nidulans* protein shown in Figure 10, and at amino acids 45-109 of the *A. niger* protein shown in Figure 28. In one embodiment, the HAC1 protein encoded by the HAC1 nucleic acid provided herein has an amino acid sequence having greater than 70% similarity to the sequence of Figure 7, Figure 8 or Figure 28. The proteins encoded by such nucleic acids are also provided herein.

In one embodiment, the nucleic acid provided herein encodes an amino acid sequence as set forth in Figure 7, Figure 8 or Figure 28. In yet another embodiment, the nucleic acid provided herein has a nucleic acid sequence as set forth in Figure 7, Figure 8 or Figure 28. The proteins encoded by such nucleic acids are also provided herein.

Further provided herein is an isolated nucleic acid encoding a PTC2 protein wherein said PTC2 protein modulates unfolded protein response and wherein said PTC2 protein has at least 70% similarity to the amino acid sequence of Figure 24 or Figure 25. In preferred embodiments the PTC2 protein has preferably at least 80%, more preferably at least 90% or more preferably at least 95% similarity to said amino acid sequences. In one aspect, the PTC2 protein has an amino acid sequence as set forth in Figure 24 or Figure 25. In another aspect, the PTC2 nucleic acid has a nucleic acid sequence as set forth in Figure 24 or Figure 25. The proteins encoded by such nucleic acids are also provided herein. It is understood that as used herein, PTC2 can be used interchangeably with ptc2 and ptcB, and that in one embodiment, PTC2 encompasses homologs. Moreover, the context in which the term is used will determine whether PTC2 is a nucleic acid, a protein or either.

Also provided herein is a nucleic acid encoding an IRE1 protein having unfolded protein response modulating activity and having at least 60% similarity to an amino acid sequence as shown in Figure 26 or Figure 27. In preferred embodiments the IRE1 protein has at least 70%, preferably at least 80%, more preferably at least 90% or even more preferably at least 95% similarity to said

amino acid sequences. In one aspect, IRE1 has an amino acid or nucleic acid sequence as shown in Figure 26 or Figure 27. It is understood that as used herein, IRE1, Ire1 and IreA can be used interchangeably, and that in one embodiment, IRE1 includes homologs. Moreover, the context in which the term is used will determine whether IRE1 is a nucleic acid, a protein or either.

The nucleic acids provided herein may be obtained from a variety of sources including but not limited to filamentous fungi such as *Trichoderma* and *Aspergillus*. In a preferred embodiment the nucleic acids are obtained from *T. reesei, A. nidulans* or *A. niger*.

Also provided herein is a cell containing a heterologous nucleic acid encoding a protein having unfolded protein response modulating activity and a heterologous nucleic acid encoding a protein of interest to be secreted. In one aspect, said protein having unfolded protein response modulating activity is selected from the group consisting of HAC1, PTC2 and IRE1. In another embodiment, said protein of interest is selected from the group consisting of lipase, cellulase, endo-glucosidase H, protease, carbohydrase, reductase, oxidase, isomerase, transferase, kinase, phosphatase, alpha-amylase, glucoamylase, lignocellulose hemicellulase, pectinase and ligninase.

Further aspects of the invention will be understood by the skilled artisan as further described below.

Brief Description of the Figures

Figure 1 depicts a map of the plasmid pMS109, an embodiment of a plasmid constructed for the expression of the truncated yeast HAC1 gene.

Figure 2 depicts a graph showing activity of α-amylase produced from yeast containing pMS109 (squares) or an empty control vector pKK1 (diamonds)

in the vertical bar, over time, horizontal bar, and further showing the activity is greater wherein pMS109 is present.

Figure 3 depicts a bar graph showing activity of invertase produced from yeast containing pMS109 (black bars) or an empty control vector pKK1 (shaded bars) in the vertical bar, over time, horizontal bar, and further showing the activity is greater wherein pMS109 is present.

Figure 4 depicts a graph showing activity of α-amylase produced from yeast wherein HAC1 has been disrupted (diamonds) or from its parental control strain (squares) in the vertical bar, over time, horizontal bar, and further showing that the activity is greater wherein HAC1 has not been disrupted.

Figure 5 depicts a graph showing activity of *Trichoderma reesei* (*T. Reesei*) endoglucanase EGI produced from yeast wherein HAC1 has been disrupted (diamonds) or from its parental control strain (squares) in the vertical bar, over time, horizontal bar, and further showing that the activity is greater wherein HAC1 has not been disrupted.

Figure 6 depicts a map of the plasmid pMS119, where the full-length *T. reesei* HAC1 cDNA without the 20 bp intron is in the pBluescript SK$^-$ vector.

Figure 7 depicts an embodiment of a nucleotide (SEQ ID No. 1) and deduced amino acid sequence (SEQ ID No. 2) of *T. reesei* HAC1. The introns are shown in lower case letters.

Figure 8 depicts an embodiment of a nucleotide (SEQ ID No. 3) and deduced amino acid sequence (SEQ ID No. 4) of *Aspergillus nidulans* (*A. nidulans*) hacA. The introns are shown in lower case letters.

Figure 9 depicts the hairpin loop structures forming at the 5' end of the 20 bp introns in the *T. reesei* HAC1 and *A. nidulans* hacA mRNAs and at the 3' end

of the intron of the *S. cerevisiae* HAC1 mRNA. The conserved nucleotides in the loop region are shown in bold. The cleavage site of the yeast intron and the three possible cleavage sites of the *T. reesei* HAC1 intron are shown by arrows. Alignment of the 20 bp intron areas of the *T. reesei* HAC1 and *A. nidulans* hacA is shown below. The intron is in lower case.

Figure 10 depicts an amino acid sequence alignment of the *T. reesei* HAC1, *A. niduans* hacA and *S. cerevisiae* HAC1 proteins. Identical amino acids are shown by asterisks and similar ones by dots. Yeast HAC1 is homologous to the other sequences at the DNA binding domain area. The DNA binding domain is approximately at amino acids 84-147 for T. reesei (SEQ ID No. 5), and approximately at amino acids 53-116 for A. nidulans (SEQ ID No. 6).

Figure 11 depicts Northern hybridization of RNA samples derived from *T. reesei* mycelia treated with DTT (+DTT) and untreated control mycelia (-DTT). The timepoints (in minutes) after DTT addition are shown. The probes used for hybridization are shown on the left.

Figure 12 depicts Northern hybridization of RNA samples derived from *A. nidulans* mycelia treated with DTT (+DTT) and untreated control mycelia (-DTT). The timepoints after DTT addition are shown. The probes are shown on the left.

Figure 13 depicts a map of the plasmid pMS131, where the full-length *T. reesei* HAC1 cDNA without the 20 bp intron is under the yeast *PGK1* promoter in the vector pAJ401.

Figure 14 depicts a map of the plasmid pMS132, where the *T. reesei* HAC1 cDNA without the 5' flanking region and without the 20 bp intron is under the yeast *PGK1* promoter in the vector pAJ401.

Figure 15 depicts complementation of *S. cerevisiae* HAC1 and IRE1 disruptions (DHAC1 and DIRE1, respectively) with different forms of the *T. reesei*

HAC1 cDNA. The growth of transformants on media with and without inositol is shown. pAJ401 is the expression vector without any insert. pMS131 has the full-length *T. reesei* HAC1 cDNA in pAJ401. pMS132 has the *T. reesei* HAC1 cDNA without its 5' flanking region in pAJ401.

Figure 16 depicts bandshift experiments, where the binding of the malE-HACI fusion protein to the putative UPR element sequences found in *T. reesei pdi1* and *bip1* promoters was tested. The oligonucleotides used in the binding reactions are shown on the top. Lanes 1, 12 and 16, no protein; lanes 2, 4-7, 8-11, 13-15 and 17-19, malE-HACI fusion protein; lane 3, malE protein alone. The binding was competed with unlabelled oligonucleotides on lanes 5 (20 x excess); lanes 6, 10, 14 and 18 (50 x excess) and lanes 7, 11, 15, and 19 (200 x excess). Alignment of the UPR element sequences that bind the HACI-malE protein is shown below.

Figure 17 depicts a graph which shows activity of α-amylase by yeast strains expressing the *T. reesei* HAC1 cDNA without the 5' flanking region and the 20 bp intron (pMS132) (squares) and control strains with the expression vector alone (pAJ401) (diamonds) in the vertical bar over time, horizontal bar, and which further shows that activity is greater wherein pMS132 is present.

Figure 18 depicts a bar graph which shows activity of invertase by yeast strains expressing the *T. reesei* HAC1 cDNA without the 5' flanking region and the 20 bp intron (pMS132) and control strains with the expression vector alone (pAJ401) in the vertical bar, over time (horizontal bar) and which further shows that activity is greater is greater wherein pMS132 is present.

Figure 19 depicts Northern hybridization of RNA samples from a yeast strain expressing the *T. reesei* HAC1 cDNA without the 5' flanking region and the 20 bp intron (pMS132) and a control strain with the expression vector alone (pAJ401). The probes used for hybridization are shown. The signals were quantified with a phosphoimager and the KAR2 signal intensities were

normalised with respect to the TDH1 signal intensities. The normalised KAR2 signals are shown on the bottom wherein it is shown that pMS132 has greater signal.

Figure 20 depicts a map of the plasmid pMS136, where the *T. reesei* HAC1 cDNA without the 5' flanking region and the 20 bp intron is under the *A. nidulans* gpdA promoter in the vector pAN52-NotI.

Figure 21 depicts Northern hybridization of RNA samples derived from transformation of the plasmid pMS136 into a *T. reesei* strain producing CBHI-chymosin fusion protein. Samples from the parental strain (lanes 1, 5 and 9), two positive transformants (lanes 2, 3, 6, 7, 10 and 11) and a HAC1 mutant strain desginated number 31 generated in the transformation (lanes 4, 8 and 12) are shown. The growth times are shown on the top and the probes used for the hybridization on the left. Quantifications of the pdi1 and bip1 signals normalised with respect to gpd1 signals are shown on the bottom.

Figure 22 depicts Northern hybridization of RNA samples derived from mycelia of the HAC1 mutant strain number 31 treated with DTT (+DTT) and untreated control mycelia (-DTT). The timepoints after DTT addition are shown on the top and the probes used for hybridization on the left. Quantifications of the pdi1 signals normalised with respect to gpd1 signals are shown on the bottom.

Figure 23 is a graph depicting production of calf chymosin by the HAC1 mutant transformant number 31 (diamonds) and its parental strain (squares) during a shake flask culture. The chymosin (CHV) units per ml of culture are shown (vertical bar) over time (horizontal bar), and it is shown that the control has more units than the mutant.

Figure 24 depicts an embodiment of a nucleotide (SEQ ID No. 7) and deduced amino acid sequence (SEQ ID No. 8) of the fragment isolated from the *A. nidulans* ptcB gene. The intron is shown in lower case.

Figure 25 depicts an embodiment of a nucleotide (SEQ ID No. 9) and deduced amino acid sequence (SEQ ID No. 10) of the *T. reesei* ptc2 cDNA.

Figure 26 depicts an embodiment of a nucleotide (SEQ ID No. 11) and deduced amino acid sequence (SEQ ID No. 12) of the fragment isolated from the *A. nidulans* ireA gene. The intron is shown in lower case.

Figures 27A-27C depict an embodiment of a nucleotide (SEQ ID No. 13) and deduced amino acid sequence (SEQ ID No. 14) of the *T. reesei* IRE1 gene. The intron is shown in lower case.

Figure 28A-28C. The nucleotide (SEQ ID No. 15) and deduced amino acid sequence (SEQ ID No. 16) of *Aspergillus niger* var. *awamori hacA* cDNA. The 20bp unconventional intron (SEQ ID No. 17) is shown in lower case letters. The amino acid sequences of the upstream open reading frame (SEQ ID No. 18) and the HACA protein (SEQ ID No. 19) are shown below the nucleotide sequence.

Figure 29. Map of the plasmid pMS152 where the *Aspergillus niger* var. *awamori hacA* without the 5' flanking region and the 20bp intron is under control of the *Aspergillus niger* var. *awamori glaA* promoter.

Figure 30.The levels of chymosin activity measured in supernatants from duplicate cultures of strain ΔAP3pUC*pyr*GRG3#11 (ctrl) and transformants (#1, #2, #3 and #4) of this strain with pMS152.

Figure 31.The levels of laccase activity measured in supernatants from duplicate cultures of strain ΔAP4:pGPTlaccase (ctrl) and transformants (#1, #2, #3, #4, #5, #6, #7 and #8) of this strain with pMS152.

Detailed Description of th Invention

The invention will now be described in detail by way of reference only using the following definitions and examples. All patents and publications, including all sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference.

Provided herein are methods and compositions for increasing the secretion of a protein in a cell comprising inducing an elevated unfolded protein response (UPR). The compositions provided herein include nucleic acids, proteins, and cells.

In one embodiment UPR refers to the unfolded protein response which occurs in response to an increase in unfolded protein in the ER. In a method provided herein, the UPR is elevated. In one embodiment, "elevated" UPR refers to an increase in the response compared to the response which would have been induced based on the amount of unfolded protein in the ER. In one embodiment, elevated refers to an increase with respect to the length of time the response occurs. In each embodiment, the elevated UPR results in an increased capacity for the cell to produce secreted proteins compared to another cell of the same type containing the same amount of unfolded protein in the ER. Preferably, the cell having an elevated UPR in accordance with the present invention produces more secreted protein in the same amount of time as a cell not having an elevated UPR.

In one aspect, the method includes inducing the elevated UPR by modulating the amount or presence of one or more UPR modulating proteins in said cell. In one embodiment, the UPR modulating protein is selected from the group consisting of HAC1, PTC2 or IRE1. UPR modulating proteins are further discussed below. It is understood that the modulating protein can be obtained by increasing the presence of a nucleic acid which encodes a modulating protein. The protein used in the methods herein have UPR modulating activity as further

discussed below, and the nucleic acids encode a protein which has UPR modulating activity. Modulating means that an increase in the protein can lead to an increase or a decrease in the UPR. Thus, in one embodiment, the presence of a modulating protein is increased as further discussed below to reach an elevated UPR. In another embodiment, the modulating protein is decreased or eliminated to reach an elevated UPR. In a preferred embodiment, HAC1 and/or IRE1 are increased so as to reach an elevated UPR.

In one embodiment, inducing UPR means that the unfolded protein response as a whole is induced or maintained as it would be by unfolded protein in the ER. The unfolded protein response involves increased expression and regulation of multiple ER foldases and chaperones. Thus, in one embodiment, manipulation of ER foldases or chaperones on an individual gene basis would not be considered an induction of UPR. Thus, in a preferred embodiment, UPR modulating activity results in an elevated UPR wherein an elevated UPR results in upregulation of ER chaperones and foldases and increased secretion of proteins.

The nucleic acids encoding the UPR modulating proteins can be obtained from a variety of sources. Preferred organisms include but are not limited to *Saccharomyces cerevisiae, Aspergillus* spp. and *Trichoderma* spp. Also other suitable yeasts and other fungi, such as *Schizosaccharomyces pombe, Kluyveromyces lactis, Pichia* spp., *Hansenula* spp., *Fusarium* spp., *Neurospora* spp. and *Penicillium* spp. can be used. Homologous genes from other organisms can also be used. In one aspect, homologous genes refer to genes which are related, but not identical, in their DNA sequence and/or perform the same function are homologous with each other and are called each other's homologues.

HAC1, PTC2, or IRE1 amino acid and nucleic acid sequences have been described for yeast. For example, for HAC1, see GenBank accession number E15694; for PTC2, see GenBank accession number U72498; for IRE1, see

GenBank accession number Z11701. Sequences of GenBank accession numbers are incorporated herein by reference. GenBank is known in the art, see, e.g., Benson, DA, et al., Nucleic Acids Research 26:1-7 (1998) and http://www.ncbi.nlm.nih.gov/. In one embodiment, HAC1, PTC2, or IRE1 are isolated from a species other than yeast, preferably a filamentous fungi, insect cell, mammalian cell or other eukaroyote. Sequences for HAC1 are provided in Figures 7, 8 and 28. Sequences for PTC2 are provided in Figures 24 and 25. Sequences for IRE1 are provided in Figures 26 and 27.

In one embodiment, the UPR modulating sequences are identified by hybridization to other nucleic acids. Additionally, sequence homology determinations can be made using algorithms.

Thus in one embodiment, the UPR modulating nucleic acid hybridizes to a complement of a nucleic acid encoding HAC1, PTC2 or IRE1. In one embodiment, the HAC1, PTC2 or IRE1 encoding sequence is selected from the sequences provided in the respective figures. In one embodiment the stringency conditions are moderate. In another embodiment, the conditions used are high stringency conditions.

"Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to reanneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature which can be used. As a result, it follows that higher relative temperatures would tend to make the reaction conditions more stringent, while lower temperatures less so. For additional details and explanation of stringency

of hybridization reactions, see Ausubel *et al.*, Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

"Stringent conditions" or "high stringency conditions", as defined herein, may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50°C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 μg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash consisting of 0.1 x SSC containing EDTA at 55°C.

"Moderately stringent conditions" may be identified as described by Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989, and include the use of washing solution and hybridization conditions (e.g., temperature, ionic strength and %SDS) less stringent that those described above. An example of moderately stringent conditions is overnight incubation at 37°C in a solution comprising: 20% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured sheared salmon sperm DNA, followed by washing the filters in 1 x SSC at about 37-50°C. The skilled artisan will recognize how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

Homologous (similar or identical) sequences can also be determined by using a "sequence comparison algorithm." Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith &

Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection.

An example of an algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul, *et al., J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. These initial neighborhood word hits act as starting points to find longer HSPs containing them. The word hits are expanded in both directions along each of the two sequences being compared for as far as the cumulative alignment score can be increased. Extension of the word hits is stopped when: the cumulative alignment score falls off by the quantity X from a maximum achieved value; the cumulative score goes to zero or below; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm then performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid

sequences would occur by chance. For example, an amino acid sequence is considered similar to a protein such as a protease if the smallest sum probability in a comparison of the test amino acid sequence to a protein such as a protease amino acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

In one embodiment, the HAC1 protein provided herein has less than 80% sequence similarity than the HAC1 yeast protein, see for example, GenBank accession number E15694, more preferably, less than 70%, more preferably, less than 60%, more preferably less than 50%, more preferably, less than 45% or 40% similarity. In another embodiment, identity is substituted for similarity.

In another embodiment, the HAC1 protein provided herein has at least 40% similarity to the amino acid sequence set forth in Figure 7 or Figure 8. More preferably, the similarity is at least 50%, more preferably, at least 60%, more preferably at least 70%, more preferably at least 80%, more preferably at least 90%, and more preferably at least 95% or 98%. In another embodiment, identity is substituted for similarity.

In another embodiment, the HAC1 protein provided herein comprises a DNA binding domain that has at least 70% similarity to the DNA binding domain set forth in Figure 10. More preferably, the similarity is at least 70%, more preferably at least 80%, more preferably at least 90%, and more preferably at least 95% or 98%. In another embodiment, identity is substituted for similarity.

As used herein, DNA binding domain refers to the domain which binds to the conserved sequence called the UPR element in promoters of genes regulated by UPR. Embodiments of a DNA binding region are shown approximately at amino acids 84-147 of the *T. reesei* protein shown in Figure 10, approximately at amino acids 53-116 of the *A. nidulans* protein shown in Figure 10 and approximately amino acids 45-109 of the A. niger protein shown in Figure

28. HAC1 homologs will have DNA binding domains which can be identified by activity or by alignment to the binding domains in Figure 10.

In one embodiment, the PTC2 protein provided herein has less than 80% sequence similarity than the PTC2 yeast protein, see for example, GenBank accession number U472498, more preferably, less than 70%, more preferably, less than 60%, more preferably less than 50%, more preferably, less than 45% or 40% similarity. In another embodiment, identity is substituted for similarity.

In another embodiment, the PTC2 protein provided herein has at least 40% similarity to the amino acid sequence set forth in Figure 24 or Figure 25. More preferably, the similarity is at least 50%, more preferably, at least 60%, more preferably at least 70%, more preferably at least 80%, more preferably at least 90%, and more preferably at least 95% or 98%. In another embodiment, identity is substituted for similarity.

In one embodiment, the IRE1 protein provided herein has less than 80% sequence similarity than the IRE1 yeast protein, see for example, GenBank accession number Z11701, more preferably, less than 70%, more preferably, less than 60%, more preferably less than 50%, more preferably, less than 45% or 40% similarity. In another embodiment, identity is substituted for similarity.

In another embodiment, the IRE1 protein provided herein has at least 40% similarity to the amino acid sequence set forth in Figure 26 or Figure 27. More preferably, the similarity is at least 50%, more preferably, at least 60%, more preferably at least 70%, more preferably at least 80%, more preferably at least 90%, and more preferably at least 95% or 98%. In another embodiment, identity is substituted for similarity.

Additionally, further homologs of the UPR modulating sequences can be identified for example by using PCR primers based on the sequences provided

herein. In yet another embodiment, naturally occurring allelic variants of the sequences provided herein may be used.

A protein has UPR modulating activity if it is able to regulate the induction of UPR. Regulate means causing an increase or decrease in the induction of the UPR. A UPR modulating protein can increase or decrease UPR induction whether or not there is a change in the amount unfolded protein in the ER. In a preferred embodiment, a UPR modulating protein has one or more of the following activities: HAC1 activity, PTC2 activity, IRE1 activity, or binds to HAC1.

Modulating the amount of or activity of the UPR modulating protein can occur by a variety of methods. For example, to increase the presence or activity of a protein in a cell, one can over-express the nucleic acid encoding the UPR modulating protein. Over-expression as used herein means that the protein encoded by the said gene is produced in increased amounts in the cell. In one embodiment, over-expression can be used interchangeably with constitutive expression or upregulation. This can be achieved by increasing the copy number of the gene by introducing extra copies of the gene into the cell on a plasmid or integrated into the genome. Over-expression can also be achieved by placing the gene under a promoter stronger than its own promoter. The amount of the protein in the cell can be varied by varying the copy number of the gene and/or the strength of the promoter used for the expression. Thus, manipulation of genes to cause induction of UPR may involve insertion into the host of multiple copies of a gene with its native promoter either on a replicating autosomal plasmid or by integration into the chromosomal DNA. It may involve fusion of the gene with a promoter region and/or transcriptional control sequences from other genes to further increase expression or to allow controlled, inducible expression. Agonists and enhancers may also be used.

In the case where it is desired to reduce the activity of a UPR modulating protein to result in elevated UPR, a number of methods may be used such as deletion of a gene or the use of antisense nucleic acids to reduce the expression

of a gene. It may involve alteration of a gene to provide a mutant form of the protein or include the use of an inhibitor of a UPR modulating protein.

In one embodiment, UPR is elevated by using a UPR inducing form of a recombinant nucleic acid encoding a UPR-modulating protein. In one embodiment, a UPR-inducing form of a recombinant nucleic acid encoding a UPR-modulating protein is a nucleic acid which has been modified to give rise to a translatable mRNA. The translatable form mimics the modified mRNA which appears in the cell on induction of UPR and which can be translated to an active UPR-modulating protein.

In one embodiment, a UPR-inducing form of a recombinant nucleic acid includes coding sequence. Coding sequence as used herein includes the nucleic acid sequence which leads to the amino acid sequence of the protein in its active form. As used herein, a nucleic acid consisting essentially of a coding sequence explicitly excludes, lacks or omits at least internal sequence which does not get translated when the active protein is encoded. Internal sequence as used herein refers to sequence which is internal to the carboxyl terminus and the amino terminus. Examples of excluded internal sequence are shown in small letters in Figures 7, 8, 24, 26, 27 and 28. The sequence may be excluded by deletion or truncation by methods known in the art.

In one embodiment a nucleic acid comprises a sequence consisting essentially of coding sequence. In this embodiment, the nucleic acid may comprise vector sequence on either side of the coding sequence but the coding sequence excludes internal sequence which does not get translated in the encoded protein's active form.

In another embodiment, a UPR modulating protein is a variant UPR modulating protein which has been varied to have increased activity. Thus in one embodiment, the activity of a UPR modulating protein is increased to elevate UPR. In one embodiment, the activity of a UPR modulating protein is increased

by maintaining the protein in its active state. For example, IRE1 is phosphorylated when the UPR pathway is turned on. Therefore, in one embodiment herein, maintaining IRE1 in its phosphorylated induces an elevated UPR.

In a preferred embodiment, IRE1 is mutated so as to constitutively have the activity of phosphorylated IRE1. In one embodiment, serine and/or threonine residues are substituted with aspartic acid to form to form an IRE1 variant having constitutive UPR inducing activity. Other substitutions to mimic a protein in its phosphorylated state are known in the art. Preferably, the mutations are performed on the nucleic acid encoding the protein.

By the term "recombinant nucleic acid" herein is meant nucleic acid, originally formed *in vitro*, in general, by the manipulation of nucleic acid by polymerases and endonucleases, in a form not normally found in nature. Generally, a nucleic acid refers to DNA, RNA or mRNA and includes a gene or gene fragment. Thus, an isolated nucleic acid, in a linear form, or an expression vector formed *in vitro* by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, i.e. using the *in vivo* cellular machinery of the host cell rather than *in vitro* manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention.

Similarly, a "recombinant protein" is a protein made using recombinant techniques, i.e. through the expression of a recombinant nucleic acid as depicted above. Generally, the term protein and peptide can be used interchangeably herein. A recombinant protein is distinguished from naturally occurring protein by at least one or more characteristics. For example, the protein may be isolated or purified away from some or all of the proteins and compounds with which it is

normally associated in its wild type host, and thus may be substantially pure. For example, an isolated protein is unaccompanied by at least some of the material with which it is normally associated in its natural state, preferably constituting at least about 0.5%, more preferably at least about 5% by weight of the total protein in a given sample. A substantially pure protein comprises at least about 75% by weight of the total protein, with at least about 80% being preferred, and at least about 90% being particularly preferred. In one embodiment, the definition includes the production of a protein from other than its host cell, or produced by a recombinant nucleic acid. Alternatively, the protein may be made at a significantly higher concentration than is normally seen, through the use of an inducible promoter or high expression promoter, such that the protein is made at increased concentration levels. Alternatively, the protein may be in a form not normally found in nature, as in the addition of an epitope tag or amino acid substitutions, insertions and deletions, as discussed below.

A recombinant cell generally refers to a cell which has been manipulated to contain a recombinant nucleic acid or protein therein.

The protein of interest to be secreted can be any protein. Wherein the protein is not naturally secreted, the nucleic acid encoding the protein may be modified to have a signal sequence in accordance with techniques known in the art. The proteins which are secreted may be endogenous proteins which are expressed naturally, but in a greater amount in accordance with the present invention, or the proteins may be heterologous. In a preferred embodiment, the proteins are heterologous. Heterologous as used herein means the protein is produced by recombinant means. Therefore, the protein may be native to the cell, but is produced, for example, by transformation with a self replicating vector containing the nucleic acid encoding the protein of interest. Alternatively, recombinant could be wherein one or more extra copies of the nucleic acid are integrated into the genome by recombinant techniques.

In another embodiment, the protein of interest is selected from the group consisting of lipase, cellulase, endo-glucosidase H, protease, carbohydrase, reductase, oxidase, isomerase, transferase, kinase, phosphatase, alpha-amylase, glucoamylase, lignocellulose hemicellulase, pectinase and ligninase. In another embodiment, the protein of interest is a therapeutic selected from the group consisting of vaccines, cytokines, receptors, antibodies, hormones, and factors including growth factors.

The cell in which the proteins are secreted is any cell having an upregulated protein response. Preferably, the host to be transformed with the genes of the invention can be any eukaryotic cell suitable for foreign or endogenous protein production, e.g., any *S. cerevisiae* yeast strain, (e.g., DBY746, BMA64-1A, AH22, S150-2B, GYPY55-15bA, vtt-a-63015) any *Trichoderma* spp. such as *T. longibrachiatum* and the *T. reesei* strains derived from the natural isolate QM6a, such as RUTC-30, RL-P37, QM9416 and VIT-D-79125, any *Kluyveromyces* spp*/. Sch. pombe, H. polymorpha, Pichia, Aspergillus, Neurospora, Yarrowia, Fusarium, Penicillium* spp. or higher eukaryotic cells.

Examples of mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51).

In an alternative embodiment, a plant cell can be utilized. In another embodiment, a baculovirus infected insect cell is utilized. The selection of the appropriate host cell is deemed to be within the skill in the art.

Transfer of the genes into these cells can be achieved, for instance, by using the conventional methods of transformation described for these organisms. General aspects of mammalian cell host system transfections have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, etc. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology, 185:527-537 (1990) and Mansour et al., Nature, 336:348-352 (1988).

The nucleic acid (e.g., cDNA, coding or genomic DNA) encoding the UPR modulating protein may be inserted into a replicable vector. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

For yeast secretion the signal sequence may be, e.g., the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* α-factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)] or other glycolytic enzymes [Hess et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

Transcription from vectors in mammalian host cells can be controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, e.g., the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding the protein in eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α-fetoprotein, and insulin).

Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the coding sequence, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding the protein. Still other methods, vectors, and host cells are described in Gething et al., Nature, 293:620-625 (1981); Mantei et al., Nature, 281:40-46 (1979); EP 117,060; and EP 117,058.

In one embodiment, the gene is cloned into a suitable expression vector, such as pKK1 or similar vectors comprising the appropriate regulatory regions depending on the selected host. For example, these regulatory regions can be obtained from yeast genes such as the *ADH1, GAL1 - GAL10, PGK1, CUP1, GAP, CYC1, PHO5*, or asparagine synthetase gene, for instance. Alternatively, also the regulatory regions of, for example, *HAC1* can be used to express the gene in *S. cerevisiae.* The plasmid carrying the gene is capable of replicating autonomously when transformed into the recipient yeast strain and is maintained stably in a single copy due to the presence of a yeast centromeric sequence. Alternatively, a multicopy replicating plasmid could be used or integration of the plasmid into the yeast genomic DNA could be provided for using methods known in the art.

In one embodiment herein, to express HAC1 cDNA, preferably truncated in *Trichoderma* the coding region of the inducing form of the *Trichoderma* HAC1 gene is coupled for instance between the *A. nidulans gpdA* promoter and terminator and the expression cassette is transformed into a *Trichoderma* strain producing for instance bovine chymosin or another foreign protein. In the truncated form, the unconventional introns are removed, as well as any remaining terminal end adjacent to said intron. An uncoventional intron is one which is present in the mRNA in the cell which is not undergoing UPR, but which is removed from the mRNA upon induction of the UPR. UPR would be thus induced constitutively. A higher level of expression which was inducible according to the carbon source used for growth of the fungus could be achieved by fusion of the inducing form of HAC1 with the promoter of the *T. reesei cbh1* gene.

For filamentous fungi the HAC1 gene is preferably integrated into the genome using methods known in the art. Suitable promoters in addition to the *gpdA* or *cbh1* promoters or promoter of the *HAC1* gene itself are for instance the other cellulase promoters, *cbh2, egl1, egl2*, or *tef1, pgk, pki*, the glucoamylase, alpha-amylase or the alcohol dehydrogenase promoter. In filamentous fungi transformation usually results in strains with varying copies of expression vector integrated into the genome (Penttilä *et al.*, 1987) and from these the strain with optimal level of truncated *HAC1* expression for growth and enhanced secretion can be screened.

It is understood that the methods provided herein may further include cultivating said recombinant host cells under conditions permitting expression of said secreted protein. The proteins can be collected and purified as desired. In a preferred embodiment, hydrolytic enzymes are secreted. In another embodiment, the secreted proteins are used in improved alcohol production or in processes where efficient hydrolysis of raw material is desired.

The following preparations and examples are given to enable those skilled in the art to more clearly understand and practice the present invention. They should not be considered as limiting the scope and/or spirit of the invention, but merely as being illustrative and representative thereof.

Example 1
Effect of expression of truncated *HAC1* in yeast

In order to cause constitutive induction of the unfolded protein response in *Saccharomyces cerevisiae*, a truncated version of the yeast *HAC1* gene was expressed from a centromeric plasmid. The truncated version does not include the intron of *HAC1* that in normal conditions prevents the translation of the mRNA. Thus the mRNA expressed from the plasmid is translated to HAC1 protein constitutively and causes a constitutive induction of the unfolded protein response. The appropriate *HAC1* gene fragment was first amplified from yeast chromosomal DNA by PCR. This fragment starts 24 bp before the translation start codon of the *HAC1* gene and ends with a translation stop codon inserted after the proline codon at amino acid position 220 of the deduced protein. The oligonucleotide primers used were: 5' ATC GCA GGA TTC CCA CCT ACG ACA ACA ACC GCC ACT 3' (forward primer) (SEQ ID No. 20) and 5' TAC AGC GGA TCC CTA TGG ATT ACG CCA ATT GTC AAG3' (reverse primer) (SEQ ID No. 21). BamHI restriction sites were included into both of the primers to facilitate cloning. The PCR reaction was carried out with the Vent DNA polymerase (New England Biolabs) in conditions recommended by the manufacturer. The PCR program used started with heating to 94° C for three minutes followed by 30 cycles with denaturation at 94° C for 45 seconds, annealing at 55° C for 45 seconds and synthesis at 72° C for one minute. The PCR product fragment of 690 bp was run in a 0.8% agarose gel and purified from the gel by the Qiaquick gel extraction kit (Qiagen) according to manufacturer=s protocol. The fragment was digested with BamHI and cloned into the BamHI site of the pZERO vector (Invitrogen) with methods known in the art. The *HAC1* fragment was released from pZERO by BamHI digestion and cloned into the BglII site of the vector pKK1

between the promoter and terminator of the yeast *PGK1* gene with methods known in the art. pKK1 contains the *LEU2* selectable marker gene and the centromere (CEN6) and ARS sequences for maintenance in yeast as a single-copy plasmid. The final expression plasmid was named pMS109 (Figure 1).

The plasmid pMS109 and the control plasmid pKK1 were transformed into a yeast strain producing *Bacillus amyloliquefaciens* α-amylase. In this strain, the expression cassette with the α-amylase coding region inserted between the yeast *ADH1* promoter and terminator had been integrated into the *TRP1* locus of the yeast strain DBY746 (α, *his3 1, leu2-3, ura3-52, trp1-289, Cyh*^{r}). Four pMS109 transformants and four strains transformed with the vector pKK1 were selected for cultivations. The cultivation medium was synthetic complete yeast medium without leucine (SC-Leu, described by Sherman 1991, Meth. Enzymol. 194, 3-21), buffered to pH 6.0 with 2% succinic acid and supplemented with 2% glucose as the carbon source. The 50 ml yeast shake flask cultures were inoculated to the initial OD600 (optical density at the wavelength of 600 nm) of 0.2 and growth was carried out for five days at 30° C and 250 RPM. Samples were taken daily for monitoring yeast growth and α-amylase production. α-amylase activity was measured with the Phadebas Amylase Test (Pharmacia) according to the instructions of the manufacturer. Yeast cell density was determined by measuring OD600 (optical density at the wavelength of 600 nm) of the culture. The α-amylase amounts produced by each of the pMS109 transformants were higher than the amounts produced by any of the pKK1 transformants. The average production level of pMS109 transformants was 70% higher in the end of the cultivation than the average of pKK1 clones (Figure 2). The growth of the pMS109 strains was slightly retarded when compared with the control.

To analyse the effect of the constitutive UPR induction to yeast invertase production, four clones transformed with pMS109 and four clones transformed with the pKK1 vector, derived from the α-amylase producing strain described above, were cultivated in the SC-Leu medium buffered to pH 6.0 with 2% succinic acid and containing 2% sucrose as the carbon source. The 50 ml shake flask

cultures were inoculated to the initial OD600 of 0.2 and grown subsequently for five days at 30 °C and 250 RPM. Yeast growth was followed by measuring the OD600 and samples were taken for invertase assays on days three, four and five. For each assay, cells were harvested by centrifugation from 1 ml of the culture. The cells were washed with 5 ml of 10 mM NaN_3 and resuspended in 0.2 M NaAc buffer, pH5.0 with 10 mM NaN_3. The invertase activity of the cells was measured by incubating them with 0.166 M sucrose in 0.2 M NaAc buffer, pH 5.0 for 6 minutes. The reaction was stopped by adding one volume of 0.5 M KPO_4, pH 7.0 and by separating the cells rapidly from the reaction mixture by filtration. The glucose formed into the reaction mixture was measured by the GOD-Perid kit (Boehringer Mannheim) according to the manufacturer's protocols. The invertase production of the pMS109 transformants was about 2 times higher than that of the pKK1 transformants in all the timepoints that were tested (Figure 3).

Example 2
Effect of disruption of *HAC1* in yeast

The yeast *HAC1* gene was disrupted by replacing it in the genome with a DNA fragment containing the G418 antibiotic resistance cassette flanked by 48 bp sequences from the 5' and 3' ends of the *HAC1* open reading frame. The G418 resistance cassette consists of the *E. coli* kanamycin resistance gene cloned between the promoter and terminator of the *Ashbya gossypii TEF* gene encoding translation elongation factor 1. The DNA fragment used in the disruption of the yeast *HAC1* was produced by PCR from the kanMX2 module (Wach et al., 1994, Yeast 10, 1793-1808) with the oligonucleotide primers 5' CCA CCT ACG ACA ACA ACC GCC ACT ATG GAA ATG ACT GAT TTT GAA CTA CTT GCC TCG TCC CCG CCG GGT CAC 3' (forward primer) (SEQ ID No. 22) and 5' AAT TAT ACC CTC TTG CGA TTG TCT TCA TGA AGT GAT GAA GAA ATC ATT GAC ACT GGA TGG CGG CGT TAG TAT CGA 3' (reverse primer) (SEQ ID No. 23). The PCR reaction was done with the Dynazyme DNA polymerase (Finnzymes) in conditions recommended by the manufacture. The PCR program started by denaturation at 94°C for 3 minutes, followed by 30 cycles of denaturation at 94°C for 45 seconds, annealing at 52°C for 30 seconds

and elongation at 72°C for 1 minute. A final elongation step of 5 minutes was performed at 72°C. The PCR product of about 1.5 kb was run in an 0.8% agarose gel and purified from the gel with the Qiaquick kit (Qiagen). The fragment was transformed into the yeast strain BMA64-1A (*a, ura3-1, trp1-Δ, leu2-3, 112, his3-11, ade2-1, can1-100*) with a method described (Gietz et al., 1992, Nucl. Acids Res. 20, 1425). The transformants were first grown over night on YPD plates (Sherman, 1991, Meth. Enzymol. 194, 3-21) and then replicated onto YPD plates with 200 µg/ml of the antibiotic G418. The transformants resistant to G418 were tested on plates with yeast mineral medium (Verduyn et al, 1992, Yeast 8, 501-517) with and without inositol. Chromosomal DNA was isolated from strains that were dependent on inositol, and Southern hybridization with the *HAC1* protein-coding region was performed with methods known in the art. The result of the hybridization showed that the *HAC1* gene had been disrupted in the strains dependent on inositol.

The effect of the *HAC1* disruption on the production of two heterologous proteins, the *Bacillus amyloliquefaciens* α-amylase (Ruohonen et al., 1987, Gene 59, 161-170) and the *Trichoderma reesei* endoglucanase EGI (Penttilä et al., 1987, Yeast 3, 175-185), was tested. The α-amylase was expressed from a multicopy plasmid with the *LEU2* marker gene , B485 (Ruohonen et al., 1991, J. Biotechnol. 39, 193-203, the plasmid is called YEpαa6 in this article), where the α-amylase gene has been cloned between the yeast *ADH1* promoter and terminator. The EGI was expressed from the plasmid pMP311 (Penttilä et al., 1987, Yeast 3, 175-185), where the endoglucanase cDNA has been cloned between the yeast *PGK1* promoter and terminator in a multicopy vector with the *LEU2* marker gene. The B485 and pMP311 plasmids were transformed into the *HAC1* disruptant and its parental strain with a described method (Gietz et al., 1992, Nucl. Acids Res. 20, 1425), and transformants were selected on SC-Leu plates (Sherman, 1991, Meth. Enymol. 194, 3-21). Three B485 transformants derived both from the *HAC1* disruptant and its parental strain were grown in 50 ml shake flask cultures in SC-Leu buffered to pH 6.0 with 2% succinic acid and supplemented with 2% glucose. The cultures were inoculated to the initial OD600

of 0.2, and growth was continued for four days at 30 °C and 250 RPM. The α-amylase activity in the culture supernatants was assayed as described in Example 1. The *HAC1* disruptant strain produced less than 10% of the α-amylase amount produced by the wild type control strain (Figure 4). To test the effect on EGI production, three pMP311 transformants derived from the *HAC1* disruptant and three transformants derived from the parental strain were grown in 50 ml of SC-Leu (Sherman, 1991, Meth. Enzymol. 194, 3-21) with 2% glucose in shaker flasks. The cultures were inoculated to the initial OD600 of 0.2, and grown at 30°C and 250 RPM for four days. Endoglucanase activity of the cultures was measured with the substrate 4-methylumbelliferyl-β-D-lactoside (Sigma). Supernatant samples were incubated at 50°C for 3 hours in a reaction mixture of 0.25 mg/ml of the substrate and 0.1 M glucose in 50 mM NaAc, pH 5.0. The reaction was stopped by adding two volumes of 1 M Na_2CO_3, and the absorbance of the mixture was measured at the wavelength of 370 nm. The production of the endoglucanase EGI of the *HAC1* disruptant was about 50% of the level produced by the parental strain (Figure 5).

Example 3
Cloning and sequence of the *Aspergillus nidulans hacA* and *Trichoderma reesei HAC1* genes

A homology search was performed against a public database (http://bioinfo.okstate.edu/pipeonline_db/anesquery.html) containing *Aspergillus nidulans* EST (expressed sequence tag) sequences with the yeast HAC1 protein sequence using the program BLAST (Altschul et al., 1990, J. Mol. Biol. 215, 403-410). The search identified one EST cDNA clone (c7a10a1.r1) which has homology to yeast HAC1p at the DNA binding domain. However, another region of the same cDNA clone, designated as EST c7a10a1.f1 in the database, had no obvious similarity with HAC1 and there was no annotation within the database to indicate similarity between the ESTs and HAC1. Therefore, it was unclear if the *A. nidulans* cDNA clone encoded a functional homolog of HAC1 or a different protein having a version of a DNA-binding motif. The region corresponding to the c7a10a1 EST cDNA was amplified by PCR from *A. nidulans* genomic DNA

isolated with methods known in the art. The sequences of the ends of the EST cDNA clone found from the databese were used to design the the 5' end primer (5' GCC ATC CTT GGT GAC TGA GCC 3') (SEQ ID No. 24) and 3' end primer (5' CAA TTG CTC GCT CTT ACA TTG AAT 3') (SEQ ID No. 25). The PCR reaction was performed as described in Example 2. The PCR product of 1.6 kb in length was run in an 0.8 % agarose gel, purified from the gel with the Qiaquick gel extraction kit (Qiagen) and cloned into the pGEM-AT vector (Promega) with methods known in the art. The whole fragment was sequenced from the resulting plasmid using internal oligonucleotide primers.

To isolate the *HAC1* cDNA from *Trichoderma reesei*, the proper hybridisation temperature for cDNA library screening were determined by genomic Southern hybridization with the genomic *hacA* fragment cloned from *A. nidulans* as a probe. The probe fragment was labelled with ^{32}P-dCTP using the Random primed DNA labelling kit (Boehringer Mannheim) as instructed by the manufacturer. The hybridization was performed as described (Sambrook et al., 1989, in Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) at 48°C, 50°C, 55°C and 60°C in a hybridization mixture containing 6xSSC, 5xDenhardt's, 0.5% SDS, 100 µg/ml herring sperm DNA (SSC is 0.15 M NaCl, 0.015 M Na-citrate, pH 7.0, 50xDenhardt's is 1% Ficoll, 1% polyvinylpyrrolidone, 1% bovine serum albumin). The filters were washed for 10 minutes at room temperature with 2xSSC, 0.1%SDS and for 30 minutes at the hybridization temperature with the same solution. The *T. reesei* cDNA library constructed into the vector λZAP (Stratagene, Stalbrand et al., 1995, Appl. Environ. Microbiol. 61, 1090-1097) was plated with the appropriate *E. coli* host strain, and the λ-DNA was lifted onto nitrocellulose filters (Schleicher & Schull) as instructed by the manufacturer. Hybridization of the filters was done for 18 hours at 55°C in the same hybridization mixture as the Southern hybridization. The filters with λ-DNA were washed for 10 minutes at room temperature with 2xSSC, 0.1%SDS and for 30 minutes at 55°C with the same solution. The λ-clones hybridizing with the probe were excised into pBluescript plasmids containing the cDNA inserts as instructed (Stratagene). The cDNA clone carrying the largest

insert (in the plasmid pMS119, Figure 6) was chosen for sequencing, and the whole sequence of its insert was determined with the help of internal sequencing primers. The genomic copy of the *T. reesei* gene was isolated by hybridization of a genomic λ-library in the vector λEMBL3 (Kaiser and Murray, 1985, in DNA Cloning: a Practical Approach , pp. 1-47, ed. Glover, IRL Press, Oxford). The library was plated with the appropriate *E. coli* host strain and λ-DNA was lifted onto nitrocellulose filters (Schleicher & Schull) as instructed by the manufacturer. The filters were hybridized at 42°C over night in a hybridization mixture containing 50% formamide, 5xDenhardt's, 5xSSPE, 0.1% SDS, 100 µg/ml herring sperm DNA and 1 µg/ml polyA-DNA (SSPE is 0.18 M NaCl, 1mM EDTA, 10 mM NaH_2PO_4, pH 7.7). The filters were washed for 10 minutes at room temperature with 2xSSC, 0.1% SDS and 30 minutes at 65°C in 0.1xSSC, 0.1% SDS. λ-DNA was isolated from clones hybridizing with the probe with a described method (Sambrook et al., 1989, in Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), and the genomic region corresponding to the *HAC1* cDNA was sequenced from this DNA with internal sequencing primers.

The sequences of the *Trichoderma reesei HAC1* and *Aspergillus nidulans hacA* genes are shown in Figures 7 and 8, respectively. Comparison of the genomic and cDNA sequences from both fungi (the cDNA sequence of *hacA* available in the EST database) reveals a conventional intron with consensus border sequences at a conserved position in both of the genes. A second intron of 20 bp is found in the *T. reesei HAC1* gene. This intron does not have the consensus 5' border sequence (GT). The sequence around the 5' end of this intron is predicted to have a strong tendency to form a RNA secondary structure called hairpin loop. The area between the stems of the loop has a sequence very similar to the consensus sequence found at both of the intron borders of the unconventional intron of 252 bp found in yeast *HAC1* (Figure 9, Gonzalez et al., 1999, EMBO. J. 18, 3119-3132). When the yeast UPR pathway is induced, the IRE1 protein cleaves the *HAC1* mRNA at these intron borders, and thus initiates the splicing of the intron and formation of an active HAC1 protein. In the

Aspergillus nidulans hacA gene there is a sequence almost identical to the hairpin-unconventional intron region of *T. reesei HAC1*.

It has been shown by RT-PCR studies that the 20 bp intron is removed from the *T. reesei HAC1* and *A. nidulans hacA* mRNAs upon UPR induction (Example 4). The 250 bp intron in yeast *HAC1* prevents translation of the mRNA probably by forming a specific secondary structure (Chapman and Walther, 1998, Curr. Biol. 7, 850-859). The 20 bp intron in the *HAC1/hacA* genes of filamentous fungi can not form such secondary structures, and thus the activation mechanism of these genes is different from yeast *HAC1*. The *T. reesei HAC1* cDNA encodes an open reading frame of 451 amino acids and the *A. nidulans hacA* a protein of 350 amino acids, when the 20 bp introns have been removed from the both sequences. The putative *T. reesei* and *A. nidulans* HACI/A proteins have an identity of 37.4% with each other and both have a DNA binding domain conserved with yeast HAC1 protein (Figure 10). The yeast HAC1 binding site has approximately 64% similarity and 53% identity to the binding site of *T. reesei*, and approximately 65% similarity and 56% identity to the binding site of *A. nidulans*. At other regions there is no detectable homology between yeast HAC1p and the HACI of *T. reesei* or the HACA of *A. nidulans*. The *HAC1* cDNA clone sequenced from *T. reesei* has a 5' flanking region of 471 bp, containing two short open reading frames encoding 17 and 2 amino acids. The 5' flanking region sequenced from *A. nidulans hacA* is 187 bp in length, containing one upstream open reading frame of 7 amino acids.

Example 4
Demonstration of truncation at the 5' end and splicing of the 20 bp intron of *T. reesei* and *A. nidulans HAC1/hacA* mRNA upon UPR induction

When the UPR pathway is induced in yeast, the unconventional intron of the *HAC1* gene is spliced and thus the length of the *HAC1* mRNA is reduced by 250 bp (Cox and Walter, 1996, Cell 87, 391-404). It was studied if UPR induction affects the length of the *HAC1/hacA* mRNA in *T. reesei* and *A. nidulans*. The *T. reesei* strain RutC-30 (Montenecourt and Eveleigh, 1979, Adv. Chem. Ser. 181,

289-301) was grown in a shake flask in a *Trichoderma* minimal medium (Penttilä et al., 1987, Gene 61, 155-164) with 2% lactose as the carbon source. Growth was performed for 60 hours at 28°C and 200 RPM, and the mycelium was diluted 1:10 into the same medium and grown for additional 21 hours. The culture was subsequently divided into two halves, and one half of it was treated with 10 mM dithiothreitol (DTT) to induce the UPR pathway (Saloheimo et al., 1999, Mol. Gen. Genet. 262, 35-45). Mycelial samples were collected from the culture treated with DTT and the untreated control culture before DTT addition and 30, 60, 90, 120 240 and 360 minutes after the addition of DTT. Total RNA was isolated from the samples with the TRIzol reagent (Gibco-BRL) according to manufacturer's protocols. RNA samples of 5 µg were treated with glyoxal and run in a 1% agarose gel in 10 mM Na-phosphate buffer, pH 7.0. Capillary blotting onto a Hybond-N nylon filter (Amersham) was done as instructed by the manufacturer. The full-length *HAC1* cDNA that was used as a probe was labelled as described in Example 3. Hybridization was performed for 18 hours at 42°C in 50% formamide, 10% dextran sulphate, 1% SDS, 1 M NaCl and 125 µg/ml of herring sperm DNA. The filter was washed in 5xSSPE for 15 minutes at 42°C, in 1xSSPE, 0.1% SDS for 30 minutes at 42°C and in 0.1xSSPE, 0.1% SDS for 30 minutes at room temperature. The results (Figure 11) show that the length of the *HAC1* mRNA does not change in the control samples not treated with DTT. In the samples treated with DTT a shorter mRNA of about 2.2 kb appears in addition to the 2.5 kb mRNA observed in the control samples. The full-length *HAC1* cDNA probe was removed from the Northern filter by incubating it in 0.1% SDS at 100°C for 10 minutes. The filter was then hybridized with a probe containing a 160 bp sequence from the 5' flanking region of the *HAC1* gene. This probe was made by PCR from the plasmid pMS119 (Figure 6) with the T3 primer (5' AAT TAA CCC TCA CTA AAG GG 3') (SEQ ID No. 26) binding to the pBluescript vector as the forward PCR primer and the oligonucleotide 5' TGG TTG ATG ACG ACG ATGCGA ACA GTC ATG ACA GGC AAC G 3' (SEQ ID No. 27) as the reverse primer. The PCR reaction was performed as described in Example 2. The probe preparation was done as in Example 3. The Northern hybridisation with the short fragment was done as described above for the full-length *HAC1*

cDNA probe. The short probe fragment derived from the 5' flanking region of the *HAC1* cDNA hybridized with the full-length *HAC1* mRNA of 2.5 kb but not with the 2.2 kb mRNA that appears when UPR is induced by DTT, indicating that the 5' end is the segment absent in the 2.2 kb mRNA. It has previously been shown that the *T. reesei pdi1* gene is controlled by the UPR (Saloheimo et al., 1999, Mol. Gen Genet. 262, 35-45). To show that the UPR is induced in this experiment with DTT, the filter was probed with the *pdi1* and *gpd1* probes. The *pdi1* mRNA becomes more abundant in the mycelium treated with DTT, whereas the *gpd1* mRNA remains at an almost constant level.

To analyse more closely the change that occurs in the *T. reesei HAC1* mRNA upon UPR pathway induction, the mRNA populations in induced and uninduced conditions were studied by rapid amplification of cDNA ends by PCR (RACE-PCR). PolyA+ RNA was isolated from total RNA samples derived from a DTT-treated and an untreated control mycelia, using the OligoTex mRNA isolation kit (Qiagen) as instructed by the manufacturer. The Marathon cDNA amplification kit (Clontech) was used in the RACE-PCR procedure according to manufacturer's protocols. The *HAC1*-specific oligonucleotide used in the reaction was 5' GGG AGA CGA CTG CTG GAA CGC CAT 3' (SEQ ID No. 28). It binds 500 bp downstream from the 5' end of the full-length *HAC1* cDNA. The isolated mRNA was used in synthesis of double-stranded cDNA. An oligonucleotide adapter was ligated to the ends of the cDNA, and the 5' ends of the *HAC1* cDNAs in each sample were amplified by PCR with the *HAC1*-specific primer and a primer supplied in the kit that binds to the ligated adapter. The PCR program consisted of 5 cycles with denaturation at 94°C for 5 seconds and synthesis at 72°C for 3 minutes followed by 5 cycles with denaturation at 94°C for 5 seconds and synthesis at 70°C for 3 minutes and 25 cycles with denaturation at 94°C for 5 seconds and synthesis at 68°C for 3 minutes. The PCR products were analysed in a 1% agarose gel. A fragment of the expected size (about 550 bp, including the 5' flanking region of the *HAC1* gene and the adapter ligated to the end), corresponding to the 2.5 kb mRNA, was obtained from the control sample derived from mycelia not treated with DTT. A second fragment of about 250 bp,

corresponding to the 2.2 kb mRNA size, was obtained in the PCR from the sample treated with DTT in addition to the one observed in the control sample. The 550 bp fragment of the control sample and the 250 bp fragment from the DTT-treated sample were isolated from the agarose gel with the Qiaquick gel extraction kit (Qiagen) as instructed by the manufacturer, and cloned into the pCR2.1-TOPO vector with the TOPO TA cloning kit (Invitrogen) as instructed by the manufacturer. Two independent clones derived from the control RNA and carrying the 550 bp insert were sequenced. They had their 5' end 8 bp and 16 bp downstream from the 5' end of the full-length cDNA (nucleotides 8 and 16 in the sequence in Figure 7) and the sequence continued until the end of the *HAC1*-specific primer as in Figure 7. Seven independent clones derived from the DTT-treated mycelium and carrying 250 bp inserts were sequenced. The 5' ends of these fragments were each at different positions between nucleotides 254 and 336 in the sequence in Figure 7, and in each case the sequence continued untill the end of the *HAC1*-specific primer as in Figure 7. This further confirms that the 5' end of the *T. reesei HAC1* mRNA is absent when the UPR pathway is induced by DTT. The upstream open reading frame (uORF) of 17 amino acids is in the region that is left out from the mRNA. Thus this uORF could be involved in the regulation, preventing translation initiation at the correct start codon and formation of the HACI protein.

The splicing of the 20 bp intron from the *T. reesei HAC1* mRNA upon UPR induction was studied by reverse transcriptase-PCR (RT-PCR). The mRNA samples used in RACE-PCR (previous paragraph), one treated with 10 mM DTT and the other not treated, were subjected to first strand cDNA synthesis with the Riboclone cDNA synthesis system (Promega) according to manufacturer's instructions. A fragment of about 500 bp in length, covering the region with the 20 bp intron in the *HAC1* gene, was amplified by PCR from the synthesized cDNA using the forward primer 5' CCC CGA GCA GTC CTT GAT GG 3' (SEQ ID No. 29) and the reverse primer 5' GTC GTT GAT GTC GAA GT 3' (SEQ ID No. 30). The PCR program consisted of denaturation at 94°C for 2 minutes followed by 30 cycles with denaturation at 94°C for 45 seconds, annealing at 50°C for 30

seconds and synthesis at 72°C for 1 minute. A final synthesis step of 5 minutes at 72°C was performed. The DNA fragments obtained in the PCR were cloned into the pCR2.1 vector with the TOPO TA cloning kit (Invitrogen) as instructed by the manufacturer. Ten fragments derived from both the DTT-treated sample and the nontreated control sample were sequenced. Nine out of the ten fragments from control sample had the intron unspliced. Only two out of the ten fragments from the DTT-treated sample had the intron unspliced, showing that splicing of the intron occurs upon UPR induction by DTT.

To examine whether the 5' flanking region and the 20 bp intron are removed from the *Aspergillus nidulans hacA* mRNA upon UPR induction similarly to the *T. reesei HAC1* mRNA, Northern hybridisation and RT-PCR experiments were carried out. The *A. nidulans* strain FGSC A26 was grown for three days in shake flasks in a medium containing 3% glucose, 2.5% corn steep liquor, 15 g/l KH_2PO_4, 5 g/l $(NH_4)_2SO_4$, 5 mg/l FeSO, 1.6 mg/l $MnSO_4$, 1.4 mg/l $ZnSO_4$, 3.7 mg/l $CoCl_2$, pH 6.8. The culture was divided into two aliquots, and one aliquot was treated with 20 mM DTT and the other served as a control. Samples were withdrawn from both aliquots at 0, 30, 60, 120 and 240 minutes after the DTT addition. The mycelium was washed with 0.9 % NaCl and stored frozen at -70°C. Total RNA was isolated from the mycelia with the Trizol reagent (Gibco-BRL) as instructed by the manufacturer. Agarose gel electrophoresis, Northern blotting and hybridization of the RNA samples was performed as described in the first paragraph of this example. The Northern was first probed with the full-length *hacA* genomic fragment shown in Figure 8. The probe hybridizes with a single 1.7 kb mRNA band in samples not treated with DTT. In the samples treated with DTT for 120 and 240 minutes, an additional band of about 1.55 kb is detected, showing that the *hacA* mRNA is truncated upon UPR induction (Figure 12). The Northern was then probed with a short probe derived from the 5' end of the *hacA* gene. The probe fragment was made by PCR from the pGEM-AT vector carrying the *hacA* gene (Example 3) with the T7 primer (5' GTA ATA CGA CTC ACT ATA GGG C 3') (SEQ ID No. 31) as the forward primer and *hacA*-specific oligonucleotide 5' TTA GGA CAG AGG CCA CGG TGT 3' (SEQ ID No. 32) as

the reverse primer. The PCR reaction was performed as described in the previous paragraph. The 5' end probe has the first 90 bp of the sequence in Figure 8. The short 5' end probe hybridizes only with the 1.7 kb mRNA, showing that the *hacA* mRNA is truncated from the 5' end when the UPR pathway is induced.

To test if the 20 bp intron is removed from the *A. nidulans hacA* gene when UPR is induced by DTT, RT-PCR was performed. The total RNA samples isolated from mycelia treated with 20 mM DTT for 240 minutes and from control mycelia were subjected to RT-PCR reactions with the Robust RT-PCR kit (Finnzymes, Finland) as instructed by the manufacturer, using the forward primer 5' CCC ATC CTT GGT GAC TGA GCC 3' (SEQ ID No. 33) and the reverse primer 5' AAG AGT CGG TGT CAG AGT TGG 3' (SEQ ID No. 34). The DNA fragment of about 400 bp obtained in the PCR was cloned into the pCR2.1-TOPO vector with the TOPO TA cloning kit (Invitrogen) as instructed by the manufacturer. Twelve of the cloned fragments derived from DTT-treated and ten from control mycelia were sequenced. None of the fragments derived from the control mycelia had the intron spliced. Three of the fragments derived from the DTT-treated mycelia had the intron spliced.

Example 5
Complementation of yeast *HAC1* and *IRE1* disruptions by different forms of the *T. reesei HAC1* cDNA

The *S. cerevisiae IRE1* gene was disrupted in the same way as the *HAC1* gene (described in Example 2). A fragment where a G418 resistance cassette is flanked by sequences from the 5' and 3' ends of the *IRE1* open reading frame was made by PCR. The forward primer 5' ATT AAT ATT TTA GCA CTT TGA AAA ATG CGT CTA CTT CGA AGA AAC ATG CTT GCC TCG TCC CCG CCG GGT CAC 3' (SEQ ID No. 35) and the reverse primer 5' AAG CAG AGG GGC ATG AAC ATG TTA TGA ATA CAA AAA TTC ACG TAA AAT GTC GAC ACT GGA TGG CGG CGT TAG TAT 3' (SEQ ID No. 36) were used in the PCR

reaction. The PCR reaction, yeast transformation, and selection and analysis of the disruptants were performed as described in Example 2 for *HAC1* disruption.

To express different forms of the *T. reesei HAC1* gene in the yeast *HAC1* and *IRE1* disruptants, four expression constructs were made into the multicopy expression vector pAJ401 (Saloheimo et al., 1994, Mol. Microbiol. 13, 11-21) with the *URA3* marker gene and yeast *PGK1* promoter and terminator to drive the expression. One of them has the *HAC1* cDNA with the intact 5' flanking region and does not have the 20 bp intron. This plasmid, pMS131 (Figure 13), was made by releasing the *HAC1* cDNA from pMS119, which is the pBluescript vector (Stratagene) carrying the full-length cDNA, with EcoRI and Asp718 digestion, filling in the ends of the fragment with Klenow polymerase and ligating it to the EcoRI restriction site of pAJ401 with methods known in the art. The second construct has the *T. reesei HAC1 cDNA* truncated at the 5' end but does not have the 20 bp intron. The truncated *HAC1* cDNA fragment was made by PCR from the plasmid pMS119 (Figure 6) with the forward primer 5' CCG CAA CAC GAC ACG GCA GGC AAC 3' (SEQ ID No. 37) and reverse primer 5' CTA GGT AGA CGT TGT ATT TTG 3' (SEQ ID No. 38). The PCR reaction was carried out as described in Example 2. The PCR product was run in a 0.8% agarose gel and purified from it with the Qiaquick gel extraction kit (Qiagen). The fragment was cloned into the EcoRV restriction site of the pZERO vector using the Zero Background Cloning kit (Invitrogen) according to manufacturer's protocols. The fragment was released from this vector with BamHI digestion and cloned between the EcoRI and XhoI restriction sites of the pAJ401 vector with methods known in the art. The resulting plasmid was named pMS132 (Figure 14). The third and fourth expression plasmids have the 20 bp intron added to the *HAC1* cDNA forms either with or without the 5' flanking region. These plasmids were constructed by replacing a HpaI-KspI fragment of about 800 bp in pMS131 and pMS132 with a corresponding HpaI-KspI fragment from a cDNA which has the 20 bp intron, isolated from the cDNA library in λZAP together with the cDNA in the plasmid pMS119 (Example 3).

To test for complementation, the four expression plasmids and the vector pAJ401 alone were transformed into the yeast *HAC1* and *IRE1* disruptants as described (Gietz et al., 1992, Nucl. Acids Res. 20, 1425). Four transformants from each of the transformations were streaked on SC-Ura plates (Sherman, 1991, Meth. Enzymol. 194, 3-21) and grown at 30 °C for three days. The plates were then replicated onto mineral medium plates (Verduyn et al., 1992, Yeast 8, 501-517) with inositol and on plates without inositol. These plates were incubated at 30 °C for three days and the streaks growing on them were replicated again onto the same plates. After growth of five days the inositol dependence of the transformants was evaluated (Figure 15). Both pMS131 (*HAC1* cDNA with 5' flanking region and without intron) and pMS132 (without 5' flanking region, without intron) could restore the ability of both the *HAC1* and *IRE1* disruptants to grow without inositol. Thus the *T. reesei HAC1* encodes the functional homolog of the yeast *HAC1* gene. When the 20 bp intron is added to pMS131, no complementation is obtained. When the intron is added to pMS132, the yeast disruptants grow very slowly without inositol. Thus the 20 bp intron weakens the ability of the *T. reesei HAC1* gene to complement the yeast *HAC1* and *IRE1* disruptions.

Example 6
Binding of the *T. reesei* HACI protein to UPR elements of the *pdi1* and *bip1* promoters

A fragment of the *T. reesei* HACI protein containing the putative DNA binding domain and leucine zipper region was produced in *E. coli* as a fusion protein with the *E. coli* maltose-binding protein malE. A DNA fragment encoding this part of the HACI protein was prepared by PCR from the *HAC1* cDNA with the oligonucleotide primers 5' TCG AAC GGA TCC GAA AAG AAG CCC GTC AAG AAG AGG 3' (forward primer) (SEQ ID No. 39) and 5' ATC GCA GGA TCC CTA GGT TTG GCC ATC CCG CGA GCC AAA 3' (reverse primer) (SEQ ID No. 40). The PCR reaction was performed as in Example 2. The PCR product of 360 bp was run in an 0.8% agarose gel and purified from the gel with the Qiaquick gel extraction kit (Qiagen). The fragment was digested with BamHI at the restriction

sites included in the PCR primers and cloned into the BamHI restriction site of the vector pMAL-p2X (New England Biolabs) with methods known in the art. The HACI-malE protein was produced in *E. coli* and purified by amylose affinity chromatography using the pMAL Protein Fusion and Purification System (New England Biolabs) as recommended by the manufacturer. The *E. coli* cells were grown up to OD600 0.5 at 37°C, IPTG was added to the concentration of 0.3 mM, and production was carried out for 3 hour at 24°C. The HACI-malE fusion protein with the expected apparent molecular weight was purified.

The oligonucleotides used in binding reactions were annealed in the concentration of 100 mg/ml in 50 mM Tris, pH 8.0, 10 mM $MgCl_2$, 1 mM spermidine and 5mM DTT by heating them at 65°C for 10 minutes and letting them cool down to room temperature during 2 hours. The oligonucleotides were labelled by incubating 100 ng of the annealed oligonucleotide in 10 mM Tris, pH 8.0, 5 mM $MgCl_2$ with 20 µCi of ^{32}P-dCTP and 2.5 U Klenow polymerase (Boehringer Mannheim) at 37°C for 30 minutes. The binding reactions between the oligonucleotides having the putative UPR elements and the proteins were carried out with 0.5 - 2 µg of the HACI-malE fusion protein or 2 µg of the malE protein and 1 ng of the annealed and labelled oligonucleotide in a mixture containing 20 mM HEPES, pH 6.9, 50 mM KCl, 10 mM $MgCl_2$, 0.25 mM EDTA, 0.5 mM DTT, 2% Ficoll, 5% glycerol and 100 µg/ml poly(dIdC) DNA. The competing oligonucleotides were used in 20-200 times excess of the labelled oligonucleotide. The binding eaction mixtures were incubated for 30 minutes at 25°C and run in a 5% polyacrylamide gel with 10% glycerol in 12.5 mM Tris-borate, pH 8.3, 0.6 mM EDTA for three hours. The gel was dried on a filter paper and exposed onto an X-ray film.

The following oligonucleotides carrying the putative UPR elements of the *pdi1* and *bip1* promoters were used in the binding reactions (only the leading strand is given, the UPREs are given in bold):

pdiUPREI+II, containing both of the putative UPR elements of the *pdi1* promoter (Saloheimo et al. 1999, Mol. Gen. Genet. 262, 35-45).
5' CGG CTG AAC CAG CGC GGC AGC CAG ATG TGG CCA AAG GG 3' (SEQ ID No. 41)

pdiUPREI, containing the UPREI of the *pdi1* promoter in a random context
5' GGT ACC TGC TAA CCA GCG CGG CAT GAT TCA AC 3' (SEQ ID No. 42)

pdiUPREII, containing the UPREII of the *pdi1* promoter in a random context
5' GGA TCT TGC ATA GCC AGA TGT GGC CTC GAT TGA CT 3' (SEQ ID No. 43)

bipUPREI, containing the UPREI of the *bip1* promoter (unpublished results)
5' GGA TTA GAA AAC GCC AAC GTG TCC ATA ACG GTC 3' (SEQ ID No. 44)

bipUPREII, containing the UPREII of the *bip1* promoter, the element is in a reverse orientation in the promoter (unpublished results)
5' GGG CGT GGA GAA GCG AGA AGT GGC CTC TTC TTC TCC 3' (SEQ ID No. 45)

The results (Figure 16) show that the HACI-malE fusion protein binds to the putative UPR element area found from the *pdi1* promoter whereas the malE protein alone does not show any binding. The binding of the fusion protein is specific, since it is competed by an excess of unlabelled oligonucleotide. The fusion protein binds specifically also to the oligonucleotide pdiUPREII and not at all to pdiUPREI, and this indicates that the functional UPR element of the *pdi1* promoter is UPREII. The HACI-malE fusion protein also binds specifically to both of the putative UPR elements found in the *bip1* promoter. Alignment of the three *T. reesei* UPR element shows that the consensus sequence for binding is GC(C/G)A(G/A)N_{1-2}GTG(G/T)C (Figure 16) (SEQ ID No. 46).

Example 7
Expression in yeast of the *Trichoderma HAC1* cDNA without its 20bp intron and truncated at the 5' end

The *T. reesei HAC1* cDNA was expressed without its 5' flanking region and without the 20 bp intron from the plasmid pMS132 (Figure 14). This plasmid and the control plasmid pAJ401 were transformed with a described method (Gietz et al., 1992, Nucleic Acids Res. 20, 1425) into the yeast strain producing *Bacillus amyloliquefaciens* α-amylase described in Example 1,. Two strains carrying pMS132 and two strains with pAJ401 were grown for six days in shake flasks (250 RPM, 30°C) in SC-Ura medium (Sherman, 1991, Meth. Enzymol. 194, 3-21) buffered to pH 6.0 with 2% succinic acid and growth and amylase production were assayed as described in Example 1. Cell samples were withdrawn from the culture for Northern analysis. The α-amylase production of the pMS132 transformants calculated per biomass was higher than that of the pAJ401 transformants from day 3 until the end of the cultivation (Figure 17). Growth of the pMS132 strains was slower than the growth of the control plasmid strains. Four pMS132 transformants and four pAJ401 transformants were grown in shake flasks (250 RPM, 30°C) in SC-Ura with 2% sucrose as the carbon source, and invertase activity produced by the cells was assayed as described in Example 1. More invertase was produced by the pMS132 transformants than by the pAJ401 transformants (Figure 18).

To show that the truncated *T. reesei HAC1* cDNA is beneficial for α-amylase and invertase production by inducing the UPR pathway of yeast, Northern analysis was performed on the cell samples withdrawn from the cultures of pMS132 and pAJ401 transformants. Total RNA was isolated from the cells collected after 1, 2 and 3 days of growth with the RNeasy RNA extraction kit (Qiagen) as instructed by the manufacturer. The yeast *KAR2* gene is under the UPR pathway control (Cox and Walter, 1996, Cell 87, 391-404), and therefore the Northern filter was probed with a fragment derived from *KAR2*. This fragment was produced by PCR from yeast chromosomal DNA with the oligonucleotide primers 5' GTG GTA ATA TTA CCT TTA CAG 3' (SEQ ID No. 47) (forward

primer) and 5' CAA TTT CAA TAC GGG TGG AC 3' (reverse primer) (SEQ ID No. 48). A fragment from the yeast *TDH1* gene encoding glyceraldehyde phosphate dehydrogenase was used as a control probe, since this gene is expressed constitutively and is not expected to be affected by UPR. The *TDH1* probe fragment was made from yeast chromosomal DNA by PCR with the oligonucleotide primers 5' TGT CAT CAC TGC TCC ATC TT 3' (forward primer) (SEQ ID No. 49) and 5' TTA AGC CTT GGC AAC ATA TT 3' (reverse primer) (SEQ ID No. 50). The PCR reaction was done as in Example 2 and the probes were prepared as described in Example 3. Northern blotting and hybridization were performed from the RNA samples as described in Example 4. The filter was exposed to the screen of the phosphoimager SI (Molecular Dynamics), and the signal intensities were quantified with the phosphoimager. The *KAR2* signal intensities were normalized with reference to the *TDH1* signal intensities. The results (Figure 19) show that the *KAR2* mRNA abundance is 2-4-fold higher in the pMS132 transformants than in the pAJ401 transformants in all the timepoints.

Example 8
Expression in *Trichoderma reesei* of the *HAC1* gene without its 20 bp intron and truncated at the 5' end

To induce the UPR pathway constitutively, a form of the *T. reesei HAC1* cDNA that is truncated at its 5' flanking region and does not have the 20 bp intron was expressed in *T. reesei*. The form of the *HAC1* cDNA that was present in pMS132 was expressed in yeast as described in Example 5 was cloned with methods known in the art into the NcoI restriction site of the vector pAN52-NotI, between the *gpdA* promoter and *trpC* terminator of *Aspergillus nidulans*. The hygromycin resistance cassette consisting of the *A. nidulans gpdA* promoter and *trpC* terminator and the *E. coli* hygromycin resistance gene was subsequently cloned into the NotI restriction site of the pAN52-NotI carrying the *HAC1* cDNA fragment. The resulting plasmid, named pMS136 (Figure 20), was transformed into *T. reesei* strain P37PΔCBHIpTEX-CHY22 as described (Penttilä et al., 1987, Gene 61, 155-164). Strain P37PΔCBHIpTEX-CHY22 was constructed by transformation of strain P37PΔCBHIPyr-26 (U.S. Patent No. 5,874,276) with a

version of the expression vector of pTEX-CHY. Vector pTEX-CHY is a derivative of pTEX in which the coding region for the *T. reesei* cellobiohydrolase I (CBHI) signal sequence, catalytic core and linker region (amino acids 1-476 of CBHI, Shoemaker, et al., 1983, Bio/Technology, 1:691-696) fused to the coding region of bovine prochymosin B (Harris et al., Nucleic Acids Research, 10:2177-2187. was inserted between the cbh1 promoter and terminator region by methods known in the art. Selection of the P37PΔCBHIpTEX-CHY22 transformants with pMS136 was performed on media with 100 µg/ml hygromycin. To obtain uninuclear transformant clones, the transformants were sporulated and single spores were plated on the selective medium with hygromycin. Purified transformants and the parental strain used in the transformation were grown in shake flasks (28°C, 200 RPM) in *Trichoderma* minimal medium (Penttilä at al., 1987, Gene 61, 155-164) supplemented with 3% whey and 0.2% peptone. Mycelial samples were collected from the cultures on the third, fifth and sixth cultivation days. Total RNA was isolated from the mycelia with the TRIzol reagent (Gibco-BRL) as instructed by the manufacturer. Northern blotting and hybridization were performed to the RNA samples as described in Example 4. The Northern filter was first probed with the full-length *HAC1* cDNA, and an mRNA derived from the expression construct which is about 2.0 kb in length can be observed in two of the transformants in addition to the 2.5 kb band that is derived from the native *HAC1* gene (Figure 21). The *HAC1* probe was removed from the Northern filter by incubating it in 0.1% SDS at 100°C for 10 minutes. The filter was subsequently probed with the *T. reesei pdi1, bip1* and *gpd1* probes. *Pdi1* encodes the protein disulphide isomerase and has been shown to be regulated by the UPR pathway (Saloheimo et al., 1999, Mol. Gen. Genet. 262, 35-45). *Bip1* (unpublished) encodes the *T. reesei* homologue of the ER-specific chaperone protein Bip. The *gpd1* gene encodes glyceraldehyde phosphate dehydrogenase and was used as the constitutive control probe. After hybridization the filter was exposed to the screen of the Phosphoimager SI (Molecular Dynamics) and the signals were quanified with the phosphoimager. The *pdi1* and *bip1* signals were normalized with respect to the *gpd1* signals. The results show that in the two transformants which express the truncated *HAC1*

mRNA the *pdi1* mRNA level is 4- and 7-fold higher than in the parental strain on the third culture day (Figure 21). This indicates that the UPR pathway can be induced constitutively in *Trichoderma reesei* by the expression of *HAC1* gene without its 20 bp intron and 5' flanking region

Example 9
The effect of a *T. reesei HAC1* mutation on heterologous protein production

A *Trichoderma reesei* strain where the *HAC1* gene is mutated was unexpectedly generated during the transformation of the plasmid pMS136 into the strain producing CBHI-chymosin fusion protein (Example 7). When analysing the transformants by Northern hybridization it was noticed that one of the transformants (number 31) produced several forms of the *HAC1* mRNA that are considerably shorter than 2 kb (Figure 21 lanes 4, 8 and 12). On the fifth and sixth day of the culture as described in Example 7 the unfolded protein response is induced in the parental strain of the transformation, presumably by the production of the heterologous protein chymosin. This is seen in the Northern analysis as appearance of a *HAC1* mRNA of about 2.2 kb (truncated at the 5' flanking region) and as the induction of the *pdi1* mRNA on days 5 and 6 (Figure 21). It has previously been shown that the production of antibody Fab fragments induces the *pdi1* gene (Saloheimo et al., 1999, Mol. Gen. Genet. 262, 35-45). In the transformant number 31 the 2.2 kb *HAC1* mRNA and the induction of the *pdi1*and *bip1* mRNAs are not detected, suggesting that the *HAC1* gene of this strain is functionally impaired. To further verify this, a DTT treatment experiment of the transformant number 31 was carried out. It was grown in shake flasks (28°C, 200 RPM) in the *Trichoderma* minimal medium (Penttilä at al., 1987, Gene 61, 155-164) with 3% whey and 0.2% peptone for three days. The culture was divided into two aliquots and one of them was treated with 10 mM dithiothreitol (DTT) and the other served as the control. Samples were taken from both aliquots at 0, 30, 60, 120 and 240 minutes after DTT addition. Total RNA was isolated from the mycelia and Northern hybridization was performed as described in Example 7. Hybridization of the Northern with the *HAC1* probe reveals that the

UPR induction by DTT is severely delayed in the transformant number 31. The *HAC1* mRNA of 2.2 kb is detected only 4 hours after DTT addition (Figure 22) and a 2-fold induction of the *pdi1* gene is also apparent in this timepoint. In a wild type strain the 2.2 kb *HAC1* mRNA appears and the *pdi1* induction takes place after 30 minutes of DTT treatment (Example 4, Figure 11).

The chymosin levels produced by the control strain and the transformant number 31 were measured daily from the media of the whey-peptone cultures described in example 7. The measurements were done from two parallel cultures with a milk clotting assay (Cunn-Coleman, et al., 1991, Bio/Technology, 9:976-981. Transformant number 31 produced roughly the same amount of chymosin as the parental strain on days 2 and 3 of the culture. On the later days the chymosin levels in the culture of the mutant strain started declining, whereas the control strain could still increase significantly the chymosin amount in its culture medium (Figure 23). The difference between the two strains is evident in the late stages of the culture, where the UPR pathway is induced in the parental strain but not in the strain number 31. This suggests that a functional *HAC1* gene and induction of the UPR pathway in the late culture stages is needed for efficient production of CBHI-chymosin fusion protein in *T. reesei*.

Example 10
Cloning and sequences of the *Aspergillus nidulans ptcB* and *Trichoderma reesei ptc2* genes

The yeast protein phosphatase encoded by the *PTC2* gene has been shown to be involved in the regulation of the UPR pathway (Welihinda et al., 1998, Mol. Cell. Biol. 18, 1967-1977). The IRE1 protein is phosphorylated when the UPR pathway is turned on (Shamu and Walter, 1996, EMBO J. 15:3928-3039), and Ptc2 dephosphorylates IRE1p and regulates the UPR negatively. A BLAST search (Altschul et al., 1990, J. Mol. Biol. 215, 403-410) was made with the yeast Ptc2 sequence against the public database containing *Aspergillus nidulans* EST cDNA sequences, and the cDNA clone i2c04a1 was found to be homologous to it within the database. The region corresponding to this cDNA was

amplified by PCR from *Aspergillus nidulans* genomic DNA with the oligonucleotides 5' TTG AAC AGC AGA TCG TTA CTG 3' (forward primer) (SEQ ID No. 51) and 5' TAT AAA GTT CGT CAA TAG TGG 3' (reverse primer) (SEQ ID No. 52). The PCR reaction was carried out as described in Example 2. The resulting PCR fragment was cloned into the pCR2.1 vector with the TOPO TA cloning kit (Invitrogen) as instructed by the manufacturer. It was sequenced with internal oligonucleotide primers (Figure 24). The optimal hybridization conditions for isolation of the *T. reesei ptc2* cDNA were determined by Southern hybridization of *T. reesei* genomic DNA with the *A. nidulans ptcB* fragment as described in Example 3. A *T. reesei* cDNA library constructed in λZAP (Stratagene, Stalbrandt et al., 1995, Appl. Environ. Microbiol. 61, 1090-1097) was screened by hybridization with the *A. nidulans ptcB* fragment as described in Example 3. The λ-clones hybridizing with the probe were excised into pBluescript plasmids with the cDNA inserts as instructed (Stratagene), and the clone having the longest insert based on restriction enzyme digestion was chosen for sequencing. The insert of this cDNA clone is 1830 bp in length, encoding an open reading frame of 438 amino acids (Figure 25). The putative *Trichoderma* PTCII protein (used interchangeably with PTC2) shows the highest identity among yeast proteins to Ptc2, 48%. It also shares 60% identity with the putative PTC2 protein from *Schizosaccharomyces pombe*. The *ptcB* fragment cloned from *Aspergillus nidulans* is 1264 in length (Figure 24). Based on homology with other Ptc2 sequences, an intron has been identified in the fragment. The deduced amino acid sequence is 89% identical to *T. reesei* PTCII protein over an area of 117 amino acids.

Example 11
Cloning and sequences of the *Aspergillus nidulans ireA* and *Trichoderma reesei IRE1* genes

A search with the program BLAST (Altschul et al., 1990, J. Mol. Biol. 215, 403-410) was made with the yeast IRE1 protein sequences against the public database containing *Aspergillus nidulans* EST cDNA sequences. The EST clone v1h01a1 was homologous to yeast IRE1 protein and include such annotation.

The region corresponding to this EST cDNA was amplified by PCR from *Aspergillus nidulans* genomic DNA with the oligonucleotides 5' CGG AGG CAA GAG TCA TAG ACG 3' (forward primer) (SEQ ID No. 53) and 5' CAA TAT ATT TCT GAA CCA GTA CG 3' (reverse primer) (SEQ ID No. 54). The PCR reaction was carried out as described in Example 2. The resulting PCR fragment was cloned into the pCR2.1 vector with the TOPO TA cloning kit (Invitrogen) as instructed by the manufacturer. It was sequenced with internal oligonucleotide primers. The fragment was used as a probe in isolation of the *T. reesei IRE1* gene. Optimal hybridization conditions were first determined with Southern hybridization of genomic *T. reesei* DNA as described in Example 3. A *T. reesei* genomic library constructed in λEMBL3 (Kaiser and Murray, 1985, in DNA Cloning: a Practical Approach , pp. 1-47, ed. Glover, IRL Press, Oxford) was then plated with the appropriate *E. coli* host strain and λ-DNA was lifted onto nitrocellulose filters (Schleicher & Schull) as instructed by the manufacturer. The filters were hybridized over night at 50°C in a mix containing 6xSSC, 5xDenhardt's, 0.5% SDS, 100 µg/ml herring sperm DNA (SSC is 0.15 M NaCl, 0.015 M Na.citrate, pH 7.0, 50xDenhardt's is 1% Ficoll, 1% polyvinylpyrrolidone, 1% bovine serum albumin). The filters were washed for 10 minutes at room temperature with 2xSSC, 0.1%SDS and for 30 minutes at 50°C with the same solution. λ-DNA was isolated from clones hybridizing with the probe with a described method (Sambrook et al., 1989).

Most of the protein-coding region of the genomic *IRE1* gene was subcloned into pBluescript SK⁻ as 2.1 kb and 2.4 kb BamHI fragments with methods known in the art. These fragments were sequenced with synthetic oligonucleotide primers. The two subclone fragments do not cover the whole open reading frame, and thus the 5' end of the chromosomal gene was sequenced from DNA isolated from the λ-clone isolated from the genomic library. An *IRE1* cDNA was isolated from a *T. reesei* library constructed in λZAP (Stratagene). The cDNA library was plated with the appropriate *E. coli* host and lifted onto nitrocellulose filters (Schleicher & Schüll) as instructed by the manufacturer. The probe fragment used in the screening was obtained by

digesting the 2.4 kb genomic subclone plasmid with BamHI and SmaI. The fragment of about 600 bp was run in a 0.8% agarose gel and isolated from the gel with the Qiaguick gel extraction kit (Qiagen) with manufacturer's instructions. The probe was labelled with ^{32}P-dCTP with the Random Primed DNA labelling kit (Boehringer Mannheim). The filters were hybridized at 42°C over night in a hybridization mixture containing 50% formamid, 5x Denhardt's, 5x SSPE, 0.1% SDS, 100 µg/ml herring sperm DNA and 1 µg/ml polyA-DNA (SSPE is 0.18 M NaCl, 1mM EDTA, 10 mM NaH_2PO_4, pH 7.7). The filters were washed for 10 minutes at room temperature with 2xSSC, 0.1% SDS and for 30 minutes at 65°C in 0.1xSSC, 0.1% SDS. λ-clones giving a hybridization signal were converted into pBluescript plasmids by *in vivo*-excision as instructed (Stratagene). The *T. reesei IRE1* cDNA was sequenced from one of the plasmids with internal oligonucleotide primers.

The area sequenced from the *T. reesei IRE1* gene is about 4.5 kb, and the open reading frame encodes a protein of 1233 amino acids (Figure 27). Comparison of the genomic and cDNA sequences revealed one intron. The *T. reesei* IREI protein starts with a predicted signal sequence of 25 amino acids. There is a putative transmembrane segment at positions 574-596 of the open reading frame. The N-terminal domain (before the transmembrane segment) presumably facing the lumen of the endoplasmic reticulum has 24% identity and 39% similarity over an area of 377 amino acids with yeast IRE1p. The C-terminal part with the kinase and RNAse domains is 42% identical and 59% similar over an area of 490 amino acids to yeast IRE1p. The cloned *A. nidulans ireA* fragment is 1570 bp in length (Figure 26). It encodes the kinase and RNAse domains of the IREA protein. Based in comparison with the yeast and *T. reesei* IRE1 sequences, an intron is identified in the sequence of the *ireA* fragment. The deduced *A. nidulans* IREA amino acid sequence has 52% identity over an area of 507 amino acids to the *T. reesei* IREI protein.

Example 12

Cloning and constitutive expression of the *Aspergillus nig r var. awamori hacA* cDNA

The *A. niger* var. *awamori hacA* cDNA was isolated by heterologous hybridisation with the cloned *Aspergillus nidulans hacA* fragment described in Example 3. A cDNA library constructed from *A. niger* var. *awamori* RNA in the plasmid pYES2 (Invitrogen) was plated as *E. coli* colonies, lifted onto nitrocellulose filters and screened by colony hybridisation as described for the isolation of the *T. reesei hac1* cDNA in Example 3. The hybridisation and the final washes were performed at 57° C. Positive colonies were found and examined by restriction analysis and sequencing of the cDNA ends. The longest cDNA was sequenced throughout its length from both strands. It is 1.68 kb long and encodes a protein of 342 amino acids (Figure 28). The encoded protein has 76% identity with *A. nidulans* HACA protein and 38% identity with *T. reesei* HACI protein. The *A. niger* var. *awamori hacA* cDNA has an upstream open reading frame encoding 44 amino acids. The region of the cDNA that, according to homology with the *T. reesei* and *A. nidulans hac1/A* genes, had a 20 bp intron was sequenced from five of the *A. niger* var. *awamori hacA* cDNA clones isolated. One of these clones did not have the 20 bp intron present, showing that the intron can be spliced out as is shown in Example 4 for the 20 bp introns of *T. reesei hac1* and *A. nidulans hacA* genes.

The UPR-induced form of the *A. niger* var. *awamori hacA* cDNA was expressed in *A. niger* var. *awamori* strains producing *Trametes versicolor* laccase or bovine preprochymosin which were constructed in the following manner. Strains ΔAP3 and ΔAP4 (described in Berka, R.M. et al., 1990, Gene 86:153-162) are equivalent strains which are deleted for the *pepA* gene (encoding the major extracellular aspartic proteinase) and which have a *pyrG* null mutation.

Strain ΔAP3 was transformed with pUC*pyr*GRG3 to create strain ΔAP3pUC*pyr*GRG3#11 which produces bovine preprochymosin. This strain secretes and accumulates active chymosin (an aspartic proteinase) in the culture

medium. The plasmid, pUC*pyr*GRG3, consists of the GRG3 expression cassette (encoding the *Aspergillus niger glaA* promoter, preprochymosin open reading frame and *glaA* terminator) obtained from pGRG3 (Cullen, D. et al., 1987, Bio/Technology 5:369-376) and the *Neurospora crassa pyr4* gene inserted into pUC19. Transformants of strain ΔAP3 with this plasmid were selected on the basis of uridine auxotrophy. Transformants were screened in liquid culture for chymosin production and strain ΔAP3pUC*pyr*GRG3#11 was chosen as the best producer.

Strain ΔAP4 was transformed with pGPT-LCC1 to create strain ΔAP4:pGPTlaccase which secretes *Trametes versicolor* laccase 1. The plasmid, pGPT-LCC1, is a derivative of plasmid pGPT*pyrG1* (described in Berka, R.M. and Barnett, C.C., 1989, Biotechnol. Adv. 7:127-154) which contains the *N. crassa pyr4* gene as fungal selectable marker and the *A. niger glaA* promoter and *A. niger* var. *awamori glaA* terminator region separated by cloning sites. To create pGPT-LCC1 the open reading frame for the *Trametes versicolor lcc1* cDNA (Ong, E. et al., 1997, Gene 196:113-119) was inserted between the *glaA* promoter and terminator regions in pGPT*pyrG1*. Transformants of strain ΔAP4 with this plasmid were selected on the basis of uridine auxotrophy. Transformants were screened in liquid culture for laccase production and strain ΔAP4:pGPTlaccase was chosen as the best producer.

For the over expression of *hacA*, the induced form of the *A. niger* var. *awamori hacA* cDNA was first created by deleting the 20 bp intron and truncating the 5' flanking region by about 150 bp, which omitted the upstream open reading frame. This was done by methods known in the art. The resulting *hacA* gene fragment was then cloned into an *A. niger* var. *awamori* expression vector with methods known in the art. In the final expression construct, pMS152 (Figure 29), the *hacA* gene fragment is between the *A. niger* var. *awamori glaA* (glucoamylase gene) promoter and terminator. The *A. nidulans amdS* gene endoding acetamidase was in the plasmid as a selection marker for fungal transformation.

The *hacA* overexpression construct (pMS152) was transformed into either *A. niger* var. *awamori* strain ΔAP3pUC*pyr*GRG3#11 or strain ΔAP4:pGPTlaccase. The transformations were performed as described in Penttila et. al., 1987, Gene 61, 155-164. The transformants were selected for the ability to grow on acetamide as the sole nitrogen source. Transformants were passaged three times on selective medium before they were sporulated and single spores were plated on the selective medium.

For Southern analysis the purified transformants and the parental strains were grown in shake flasks (28°C, 200 rpm) in Clofine special medium (described in WO 98/31821). Mycelial samples for total-DNA isolations were collected on the third cultivation day. The isolations were done with the DNA EASY kit (Invitrogen) according to the manufacturer's instructions. 5µg of the total DNA was cut with restriction enzyme *Hin*dIII and *Xho*I to obtain a 5.2 kb-fragment from the integrated pMS152 to indicate which transformants have the hacA overexpression cassette and samples were run in 1% agarose gel in 1xTBE-buffer. The treatment of the gels and capillary blotting onto a Hybond-N nylon filter (Amersham) were done as instructed by the manufacturer. A fragment of the *A. niger* var. *awamori* hacA cDNA labeled as described in Example 3 was used as a probe in the Southern hybridisation. The filters were hybridised at 42°C over night in a hybridisation mixture containing 50% formamide, 5xDenhart's, 5xSSPE, 0.1% SDS, 100 µg/ml herring sperm DNA and 1µg/ml poly (A)-DNA. Filters were washed as described in Example 4. A band of the expected size was obtained from all the transformants that were analysed, but not from the parental strains. This indicated that the obtained transformants were stable and that they contained intact *hacA* overexpression cassette.

Eight transformants from the laccase-producing strain and four transformants from the chymosin-producing strain shown to contain the *hacA* overexpression cassette were cultivated again for Northern analysis and measurement of the enzymatic activities. The pMS152 transformants of the strain

producing preprochymosin and the untransformed parental strain (ΔAP3pUC*pyr*GRG3#11) were cultivated in Clofine special medium in shake flasks (28°C, 200 rpm) in two parallel cultures for six days. Mycelial samples for RNA isolations were taken on the third day of the cultivation. The pMS152 transformants of the strain producing *Trametes* laccase and the untransformed parental strain (ΔAP4:pGPTlaccase) were cultivated in 8 g/litre Bacto Soytone (Difco), 12 g/litre Tryptone peptone (Difco), 15 g/litre $(NH_4)_2SO_4$, 12.1 g/litre $NaH_2PO_4.H_2O$ and 3,3 g/litre $Na_2HPO_4.7H_2O$. After autoclaving 5 ml/litre of 20% $MgSO_4$ solution, 2 ml/litre of Cu/citrate solution (110 g/litre citrate*H_2O, 125 g/litre $CuSO_4.5H_2O$), 1 ml/litre Tween 80, 300 ml/litre 50% maltose solution and 200 ml/litre of 100 mg/litre arginine was added to the medium. The cultivations were done in shake flasks (28°C, 200 rpm) in two parallel cultures for ten days. The mycelial samples for RNA isolations were taken on the second day of the cultivation. Total RNA's were isolated from all the mycelial samples using the TRIZOL reagent (Gibco-BRL) as instructed by the manufacturer. RNA samples of 5 µg were treated with glyoxal and run in 1% agarose gel in 10 mM Na-phosphate buffer, pH 7.0. Northern blottings and hybridizations were done as described in Example 4. A fragment of the *A. niger* var. *awamori hacA* cDNA labeled as described in Example 3 was used as a probe. An mRNA of the expected size from the *hacA* overexpression cassette of about 1.6 kb was observed in all the transformants studied in addition to the band of about 1.7 kb that is derived from the native *hacA* gene and that is also seen in the controls. This indicates that the 5'-truncated and intronless *hacA* coming from the overexpression cassette is expressed in the transformants.

Example 13
The effect of *A. niger* var.*awamori hacA* overexpression on heterologous protein production

Samples from the culture supernatants of the pMS152 transformants of the strain producing preprochymosin and the untransformed parental strain (ΔAP3pUC*pyr*GRG3#11) were taken on the fifth day of cultivation. The chymosin production levels were measured with a milk-clotting assay. The samples were

diluted into buffer containing 10 g/litre sodium acetate and 5 ml/litre 1M acetic acid. 200 µl of the diluted sample was added to 5ml of buffer containing 55 g/500 ml skim milk (Difco) at 30°C. The clotting of the milk was observed visually and the time that the clotting of the milk took was recorded and correlated to a known standard. All the four transformants produced 1.3 - 2.8 fold more chymosin than the parental strain (Figure 30).

Samples from the culture supernatants of the pMS152 transformants of the strain producing *Trametes* laccase and the untransformed parental strain (ΔAP4:pGPTlaccase) were taken on the fifth and seventh day of the cultivation. The laccase activity measurements were made from the supernatants and the results showed that all the transformants produce more laccase than the parental strain. Laccase activity was measured according to Niku-Paavola et al. (Niku-Paavola M-L, Karhunen E, Salola P, Raunio V (1988) Ligninolytic enzymes of the white-rot fungus Phlebia radiata. Biochem. J. 254: 877-884) using ABTS (Boehringer Mannheim; Mannheim, Germany) as a substrate. The production levels of the transformants in the fifth day samples were 3 to 7.6 fold higher than in the parental strain. On the seventh day of cultivation the transformants produced 2 to 5.4 fold more laccase than the parental strain (Figure 31).

These results demonstrate that overexpression of an inducing form of *hacA* enables production of higher levels of secreted heterologous proteins in *A. niger*.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.